CANADIAN JOURNAL OF MATHEMATICS

Journal Canadien de Mathématiques

VOL. VII - NO. 3
1955

Published for
THE CANADIAN MATHEMATICAL CONGRESS
by the
University of Toronto Press

The chief languages of the *Journal* are English and French.

Manuscripts for publication in the *Journal* should be sent to the *Editor-in-Chief*, H. S. M. Coxeter, University of Toronto. Everything possible should be done to lighten the task of the reader; the notation and reference system should be carefully thought out. Every paper should contain an introduction summarizing the results as far as possible in such a way as to be understood by the non-expert.

All other correspondence should be addressed to the *Managing Editor*, G. de B. Robinson, University of Toronto.

The *Journal* is published quarterly. Subscriptions should be sent to the *Managing Editor*. The price per volume of four numbers is $8.00. This is reduced to $4.00 for individual members of recognized Mathematical Societies.

The Canadian Mathematical Congress gratefully acknowledges the assistance of the following towards the cost of publishing this Journal:

University of British Columbia

Carleton College	Ecole Polytechnique
Université Laval	Loyola College
University of Manitoba	McGill University
McMaster University	Université de Montréal
Queen's University	Royal Military College
St. Mary's University	University of Toronto

National Research Council of Canada
and the
American Mathematical Society

AUTHORIZED AS SECOND CLASS MAIL, POST OFFICE DEPARTMENT, OTTAWA

WEAK COMPACTNESS AND VECTOR MEASURES

R. G. BARTLE, N. DUNFORD AND J. SCHWARTZ

Introduction. It is the purpose of this paper to develop a Lebesgue theory of integration of scalar functions with respect to a countably additive measure whose values lie in a Banach space. The class of integrable functions reduces to the ordinary space of Lebesgue integrable functions if the measure is scalar valued. Convergence theorems of the Vitali and Lebesgue type are valid in the general situation. The desirability of such a theory is indicated by recent developments in spectral theory.

In §1 two criteria for the conditional weak compactness of subsets of the Banach space of countably additive measures on a σ-field Σ are derived. Their force is sufficient to allow us to conclude, in §2, that if μ is a countably additive measure on Σ with values in a Banach space, then there exists a positive scalar measure ν on Σ with respect to which μ is ν-continuous (i.e., ν-absolutely continuous). This permits the development of the integration theory.

As an example of an elementary application of the integration theory we give, in §3, the "vector" generalization of the celebrated Riesz theorem on the representation of linear functionals on a space of continuous functions. This yields representation theorems for the general, the weakly compact, and the compact operators on a space of continuous functions. Some of these results are related to theorems of Gelfand **(6)** and Grothendieck **(7)**.

Our notation and terminology are standard. We mention specifically that the weak topology of a Banach space $\mathfrak{X}$ is the topology induced by its adjoint space $\mathfrak{X}^*$ in the familiar fashion, even though $\mathfrak{X}$ itself may be the adjoint of some other space. By the $\mathfrak{X}$ topology of $\mathfrak{X}^*$ we mean the topology on $\mathfrak{X}^*$ which has as a typical neighbourhood of the origin, the set $\{x^* \in \mathfrak{X}^* \mid |x^*(x_i)| < 1, i = 1, \ldots, n\}$.

1. Weakly compact sets of measures. Let S be an abstract set and Σ be a σ-field (i.e., σ-algebra) of subsets of S; sets in Σ will frequently be called the *measurable* subsets of S.

1.1. DEFINITION. By $ca(\Sigma)$ we denote the set of all countably additive real or complex valued measures defined on Σ having finite variation. A generic element of $ca(\Sigma)$ will ordinarily be denoted by the letter λ; positive measures by ν. The symbol $|\lambda|(E)$ denotes the total variation of λ over the set $E \in \Sigma$. The symbol $|\lambda|$ represents the total variation of λ over the entire space S.

Received June 20, 1954; in revised form February 20, 1955. The research contained in this paper was done under contract NONR609(04) with the Office of Naval Research.

With $|\lambda|$ as the norm, $ca(\Sigma)$ forms a real or complex Banach space. It is the purpose of this section to characterize the subsets of $ca(\Sigma)$ which are conditionally weakly compact. We shall make free use of the theorem of Eberlein **(5)** asserting that such sets K are precisely those which are weakly sequentially compact in the sense that an arbitrary sequence from K has a subsequence which converges weakly to an element of $ca(\Sigma)$. We observe that if a sequence $\{\lambda_n\}$ converges weakly, then it is bounded and $\{\lambda_n(E)\}$ is a convergent set of scalars for each $E \in \Sigma$.

Finally, we mention that we will make essential use of the Vitali-Hahn-Saks theorem **(15, p. 967)**: Let ν be a finite positive measure on Σ, let $\{\lambda_n\} \subset ca(\Sigma)$ be such that $\lim \lambda_n(E)$ exists for each $E \in \Sigma$, and let

$$\lim_{\nu(E)\to 0} \lambda_n(E) = 0$$

for each n. It follows that this last limit exists uniformly in n, and if λ is the set function defined by

$$\lambda(E) = \lim_{n\to\infty} \lambda_n(E)$$

for $E \in \Sigma$, then $\lambda \in ca(\Sigma)$.

1.2. LEMMA. *If $\{\lambda_i\}$ is a sequence in $ca(\Sigma)$, there is a positive measure $\nu \in ca(\Sigma)$ such that*

$$\lim_{\nu(E)\to 0} \lambda_i(E) = 0, \qquad i = 1, 2, \ldots.$$

Proof. The measure defined by

$$\nu(E) = \sum_{i=1}^{\infty} 2^{-i} \frac{|\lambda_i|(E)}{1 + |\lambda_i|}, \qquad E \in \Sigma,$$

has the property that if $\nu(E) = 0$ then $|\lambda_i|(E) = 0$. The conclusion then follows **(8, p. 125)**.

1.3. THEOREM.[1] *For a set $K \subset ca(\Sigma)$ to be conditionally weakly compact it is necessary and sufficient that*

(1) *the set K is bounded, and*

(2) *if $\{E_i\}$ is a sequence in Σ which decreases to the void set, then*

$$\lim_{i\to\infty} \lambda(E_i) = 0,$$

uniformly for $\lambda \in K$.

Proof. The necessity of (1) is a consequence of the uniform boundedness theorem. To show that (2) is necessary we proceed indirectly, and suppose that there exists a positive number ϵ, a sequence $\{E_i\} \subset \Sigma$ with $E_i \downarrow \phi$, and a sequence $\{\lambda_i\} \subseteq K$ such that

$$|\lambda_i(E_i)| > \epsilon, \qquad i = 1, 2, \ldots.$$

[1] The reader should observe the similarity between this theorem and some of the criteria, established by Grothendieck **(7, p. 146)**, for weak compactness in the space of bounded regular Radon) measures.

Since K is assumed to be weakly compact, we may suppose that $\{\lambda_i\}$ is weakly convergent so that $\lim \lambda_i(E)$ exists for all $E \in \Sigma$. Construct $\nu \in ca(\Sigma)$ as in the lemma; by the Vitali-Hahn-Saks theorem

$$\lim_{\nu(E)\to 0} \lambda_i(E) = 0, \qquad \text{uniformly for } i = 1, 2, \ldots.$$

Since $E_i \downarrow \phi$, it follows that $\nu(E_i) \to 0$ and hence $\lambda_i(E_i) \to 0$, which contradicts the supposition and establishes the necessity of (2).

To prove the sufficiency, let $\{\lambda_m\}$ be an arbitrary sequence in K. Construct a positive measure $\nu \in ca(\Sigma)$ as in the lemma. Let $ca(\Sigma; \nu)$ be the closed linear manifold in $ca(\Sigma)$ consisting of measures λ which are ν-continuous, i.e. such that

$$\lim_{\nu(E)\to 0} \lambda(E) = 0.$$

By the Radon-Nikodmý theorem (8, p. 128) there is an isometric isomorphism between the space $ca(\Sigma, \nu)$ and the space $L(S, \Sigma, \nu)$ of ν-integrable functions, the correspondence $\lambda \leftrightarrow f$ being given by the formula

$$\lambda(E) = \int_E f(s)\, \nu(ds), \qquad E \in \Sigma.$$

Denote the functions corresponding to the $\{\lambda_m\}$ by $\{f_m\}$. Let $\{G_n | n = 1, 2, \ldots\}$ be a basis for the open sets in the scalar field and let $E_{mn} = f_m^{-1}(G_n)$.

Let Σ_1 be the σ-field generated by the sets $\{E_{mn}\}$. Observe that each f_m is measurable with respect to Σ_1 and thus is in the space $L(S, \Sigma_1, \nu)$. Let Σ_2 be the field generated by the sets $\{E_{mn}\}$; it is known (8, p. 23) that Σ_2 consists of a countable number of sets. By a diagonal process, pick a subsequence

$$\{\lambda_{m_k}\}$$

such that

$$\lim_k \lambda_{m_k}(E)$$

exists for all $E \in \Sigma_2$. Now let Σ_3 be the class of all subsets in Σ_1 for which

$$\lim_k \lambda_{m_k}(E)$$

exists. Evidently, $\Sigma_2 \subseteq \Sigma_3 \subseteq \Sigma_1$.

We wish to show that $\Sigma_3 = \Sigma_1$ and will accomplish this by demonstrating that Σ_3 is a monotone class (8, p. 27). Let $\{E_i\} \subset \Sigma_3$ and suppose that $E_i \uparrow E$. Then $E - E_i \downarrow \phi$, so by condition (2) we have that

$$\lim_i \lambda_{m_k}(E_i) = \lambda_{m_k}(E)$$

uniformly with respect to $k = 1, 2, \ldots$. But since $E_i \in \Sigma_3$,

$$\lim_k \lambda_{m_k}(E_i)$$

exists for $i = 1, 2, \ldots$. It follows that

$$\lim_k \lambda_{m_k}(E)$$

exists and hence $E \in \Sigma_3$. Thus Σ_3 is a monotone class and $\Sigma_3 = \Sigma_1$.

We now prove that

$$\{\lambda_{m_k}\}$$

is a weakly convergent sequence in $\mathrm{ca}(\Sigma)$. By the Hahn-Banach theorem (8*a*, p. 19) it is enough to prove that this sequence is weakly convergent in the space $\mathrm{ca}(\Sigma; \nu)$, and in view of the isometric isomorphism between this space and $L(S, \Sigma, \nu)$ it suffices to show that

$$\{f_{m_k}\}$$

is a weakly convergent sequence in $L(S, \Sigma, \nu)$ or even in the subspace $L(S, \Sigma_1, \nu)$. Since this sequence is bounded it is enough to prove convergence on a fundamental set of continuous linear functionals on $L(S, \Sigma_1, \nu)$, such as the characteristic functions of subsets of Σ_1. But the convergence on this fundamental set is equivalent to the convergence of

$$\{\lambda_{m_k}\}$$

on Σ_1, which has already been established. This proves the sufficiency of the conditions.

We shall also need the following criterion:

1.4. THEOREM. *For a set $K \subset \mathrm{ca}(\Sigma)$ to be conditionally weakly compact it is necessary and sufficient that*

(1) *the set K is bounded, and*

(2′) *there exists a positive $\nu \in \mathrm{ca}(\Sigma)$ such that*

$$\lim_{\nu(E)\to 0} \lambda(E) = 0,$$

uniformly for $\lambda \in K$.

Proof. To prove the necessity of (2′), we shall show that for any positive ϵ, there is a positive $\delta(\epsilon)$ and a finite set $\{\lambda_1, \ldots, \lambda_p\} \subseteq K$ such that if $|\lambda_i|(E) < \delta(\epsilon)$ for $i = 1, \ldots, p$ then $|\lambda(E)| < \epsilon$ for $\lambda \in K$. Suppose that this statement is false for some ϵ. Let $\lambda_1 \in K$ be arbitrary; then there exists a set $E_1 \in \Sigma$ and a $\lambda_2 \in K$ such that

$$|\lambda_1|(E_1) < 2^{-1}, \quad |\lambda_2(E_1)| > \epsilon.$$

By induction, construct sequences $\{\lambda_i\} \subseteq K$, and $\{E_i\} \subseteq \Sigma$ such that

$$(*) \qquad |\lambda_i|(E_j) < 2^{-j}, \qquad 1 \leqslant i \leqslant j < \infty,$$
$$|\lambda_{j+1}(E_j)| > \epsilon, \qquad 1 \leqslant j < \infty.$$

We shall suppose that the sequence $\{\lambda_i\}$ converges weakly in $\mathrm{ca}(\Sigma)$. Construct a positive measure $\nu' \in \mathrm{ca}(\Sigma)$ by Lemma 1.2. By the Vitali-Hahn-Saks theorem

$$\lim_{\nu'(E)\to 0} \lambda_i(E) = 0, \qquad \text{uniformly for } i = 1, 2, \ldots.$$

From the construction of ν' and (*) we have that $\nu'(E_j) \leqslant 2.2^{-j}$. On the other hand, $|\lambda_{j+1}(E_j)| > \epsilon$, which is a contradiction and establishes the validity of the statement. Now let ϵ take on the values 2^{-n}, $n = 1,2, \ldots$, and let $\delta(2^{-n})$

be the corresponding numbers and $\{\lambda_{np} | 1 \leqslant p \leqslant P_n, 1 \leqslant n < \infty\} \subseteq K$ the measures whose existence has been assured. Let ν be the positive measure in $\mathrm{ca}(\Sigma)$ defined by

$$\nu(E) = \sum_{n=1}^{\infty} 2^{-n} \sum_{p=1}^{P_n} \frac{1}{P_n} \frac{|\lambda_{np}|(E)}{1 + |\lambda_{np}|}, \qquad E \in \Sigma.$$

Then if $\nu(E) < \delta(2^{-n})\{2^n P_n(1 + M)\}^{-1}$, where $M = \sup_{\lambda \in K} |\lambda|$, it is evident that $|\lambda_{np}|(E) < \delta(2^{-n})$ $(p = 1, \ldots, P_n)$ and hence $|\lambda(E)| < 2^{-n}$ for all $\lambda \in K$.

To prove the sufficiency of the conditions, we observe that (2′) implies condition (2) of the preceding theorem, so the proof of the theorem is complete. The preceding proof establishes the partial converse:

1.5. COROLLARY. *If K is a conditionally weakly compact subset of $\mathrm{ca}(\Sigma)$, then there exists a positive $\nu \in \mathrm{ca}(\Sigma)$ satisfying (2′) and such that $\nu(E) < \epsilon$ whenever $E \in \Sigma$ is such that $|\lambda|(E) < \epsilon$ for all $\lambda \in K$.*

2. Integration with respect to a vector measure. Let $\mathfrak{X}$ be a real or complex Banach space, and $\mathfrak{X}^*$ its conjugate space. We shall be concerned with a fixed additive set function μ defined on a σ-field Σ of subsets of a set S and taking values in $\mathfrak{X}$.

2.1. DEFINITION. The additive set function $\mu : \Sigma \to \mathfrak{X}$ is called a *vector measure* if $x^*\mu \in \mathrm{ca}(\Sigma)$ for every $x^* \in \mathfrak{X}^*$; thus the set functions $\{x^*\mu | x^* \in \mathfrak{X}^*\}$ on Σ to the scalar field are finite valued and countably additive.

If μ is a vector measure, we define a non-negative set function $||\mu||$ on Σ, called the *semi-variation* of μ, by

$$||\mu||(E) = \sup |\sum_{i=1}^{n} \alpha_i \mu(E_i)|, \qquad E \in \Sigma,$$

where the supremum is taken over all finite collections of scalars with $|\alpha_i| \leqslant 1$ and all partitions of E into a finite number of disjoint measurable sets. Evidently $||\mu||(E) \leqslant ||\mu||(F)$ if $E \subseteq F$; further,

$$||\mu||\Big(\bigcup_{i=1}^{\infty} E_i\Big) \leqslant \sum_{i=1}^{\infty} ||\mu||(E_i)$$

for every sequence of measurable sets $\{E_i\}$, and the inequality may be strict.

2.2. LEMMA. *If $\mu : \Sigma \to \mathfrak{X}$ is a vector measure and if $\{E_i\}$ is a sequence of disjoint measurable sets, then*

$$\mu\Big(\bigcup_{i=1}^{\infty} E_i\Big) = \sum_{i=1}^{\infty} \mu(E_i),$$

where the series converges unconditionally in the norm of $\mathfrak{X}$. The semi-variation of μ on S is finite; in fact, if $E \in \Sigma$, then

$$||\mu||(E) \leqslant 4 \sup\{|\mu(F)| \,\big|\, F \in \Sigma, F \subseteq E\} < \infty.$$

Proof. The first fact is due to Pettis **(11; 12)**. Given $x^* \in \mathfrak{X}^*$, $x^*\mu$ is a finite valued measure, so $\sup\{|x^*\mu(F)| \mid F \in \Sigma, F \subseteq E\}$ is finite for each $x^* \in \mathfrak{X}^*$. By the Uniform Boundedness theorem **(8a**, p. 26**)**,

$$\sup\{|\mu(F)| \mid F \in \Sigma, F \subseteq E\} < \infty.$$

On the other hand,

$$||\mu||(E) = \sup_{|x^*| \leq 1} |x^*\mu|(E) \leq 4 \sup_{|x^*| \leq 1} \sup_{F \subseteq E} |x^*\mu(F)|,$$

so the conclusion follows from the Hahn-Banach theorem **(8a**, p. 19**)**.

The next lemma and its corollary are crucial for our purposes.

2.3. LEMMA. *The set of numerical measures* $\{x^*\mu \mid x^* \in \mathfrak{X}^*, |x^*| \leq 1\}$ *is conditionally weakly compact as a subset of* $\mathrm{ca}(\Sigma)$.

Proof. Let $\{E_i\}$ be a decreasing sequence of sets in Σ with void intersection. Since μ is countably additive in the strong topology, $\lim \mu(E_i) = 0$ in this topology. Consequently, $\lim x^*\mu(E_i) = 0$ uniformly for $|x^*| \leq 1$. By Theorem 1.3 the desired result follows.

2.4. COROLLARY. *If* μ *is a vector measure, then there is a positive measure* $\nu \in \mathrm{ca}(\Sigma)$ *such that*

$$\lim_{\nu(E)\to 0} \mu(E) = 0, \quad \lim_{\nu(E)\to 0} ||\mu||(E) = 0.$$

The measure ν *may be chosen such that* $\nu(E) < \epsilon$ *whenever* $||\mu||(E) < \epsilon$.

Proof. The first equation follows from Lemma 2.3, Theorem 1.4, and the Hahn-Banach theorem; the second follows from the first and Lemma 2.2. The final statement is a consequence of Corollary 1.5.

We now proceed to develop a theory of integration of scalar functions with respect to the vector measure μ. A *μ-null set* is a set $E \in \Sigma$ for which $||\mu||(E) = 0$; the term *μ-almost everywhere* refers to the complement of a μ-null set. From the countable sub-additivity of $||\mu||$ it follows that the union of a countable number of μ-null sets is also a μ-null set. A scalar function f defined on S is *measurable* if for every Borel set B of scalars, $f^{-1}(B)$ is an element of Σ. A function is *simple* if it is a finite linear combination of characteristic functions of measurable sets.

If f is the simple function

$$\sum_{i=1}^{n} \alpha_i \chi_{E_i}, \qquad E_i \in \Sigma,$$

then we define the *integral* of f over a set $E \in \Sigma$ by the equation

$$\int_E f(s)\, \mu(ds) = \sum_{i=1}^{n} \alpha_i \mu(E \cap E_i).$$

The integral of a simple function is independent of the representation of the function as a linear combination of characteristic functions. Obviously, integration of simple functions over E is a linear operation. Also, the integral of a simple function is a countably additive set function with values in $\mathfrak{X}$. If f is a simple function and if E is in Σ then evidently

$$\left| \int_E f(s)\, \mu(ds) \right| \leq \sup_{s \in E} |f(s)| \cdot ||\mu||(E).$$

If f is an arbitrary measurable function, we define the *μ-essential supremum* of f on E to be the infimum of those numbers A for which $\{s \in E \mid |f(s)| > A\}$ is a μ-null set. If

$$\mu\text{-ess sup}_{s \in E} |f(s)| < \infty,$$

then we say that f is *μ-essentially bounded* on the set $E \in \Sigma$.

2.5. DEFINITION. A measurable function f is said to be *μ-integrable* if there exists a sequence $\{f_n\}$ of simple functions such that

(1) $f_n(s) \to f(s)$ μ-almost everywhere, and

(2) the sequence

$$\left\{ \int_E f_n(s)\, \mu(ds) \right\}$$

converges in the norm of $\mathfrak{X}$ for each $E \in \Sigma$.

The limit of this sequence of integrals is defined to be the *integral of f with respect to μ over the set $E \in \Sigma$*, in symbols

$$\int_E f(s)\, \mu(ds).$$

2.6. THEOREM. (a) *If $E \in \Sigma$ and f is μ-integrable, the integral of f with respect to μ over E is an unambiguously defined element of $\mathfrak{X}$;*

(b) *the integrable functions form a linear space and for E in Σ the integral $\int_E f(s)\, \mu(ds)$ is a linear map of this space into $\mathfrak{X}$:*

(c) *if f is a measurable function which is μ-essentially bounded on E, then f is μ-integrable and*

$$\left| \int_E f(s)\, \mu(ds) \right| \leq \{\mu\text{-ess sup}_{s \in E} |f(s)|\}\{||\mu||(E)\};$$

(d) *if f is μ-integrable, then $\int_E f(s)\, \mu(ds)$ is a countable additive function on Σ to $\mathfrak{X}$;*

(e) *if f is μ-integrable, then*

$$\lim_{||\mu||(E) \to 0} \int_E f(s)\, \mu(ds) = 0;$$

(f) *if U is a bounded linear operator from $\mathfrak{X}$ into a second Banach space $\mathfrak{Y}$, then $U\mu$ is a vector measure from Σ to $\mathfrak{Y}$ and for any μ-integrable function f and $E \in \Sigma$, we have*

$$U\left\{ \int_E f(s)\, \mu(ds) \right\} = \int_E f(s)\, U\mu(ds).$$

Proof. To prove (a), let $\{f_n\}$ and $\{g_n\}$ be two sequences of simple functions as in Definition 2.5; we are required to show that the two sequences of integrals approach the same limit. We define $h_n(s) = 0$, if s is a point at which either $\{f_n(s)\}$ or $\{g_n(s)\}$ fails to converge to $f(s)$, and set $h_n(s) = f_n(s) - g_n(s)$ otherwise. It is evident that $\{h_n\}$ converges to zero everywhere, and that $\{\int_E h_n(s)\,\mu(ds)\}$ converges in the norm of $\mathfrak{X}$ for $E \in \Sigma$. We must show that this sequence of integrals converges to the zero element of $\mathfrak{X}$.

Let ν be a positive measure described in Corollary 2.4. Clearly, since each h_n is a simple function,

$$(*) \qquad \lim_{\nu(E)\to 0} \int_E h_n(s)\,\mu(ds) = 0, \qquad n = 1, 2, \ldots;$$

further, the sequence of integrals $\{\int_E h_n(s)\,\mu(ds)\}$ converges for each $E \in \Sigma$, so by the Vitali-Hahn-Saks theorem,[2] the limit in (*) is uniform in n. Consequently, for each $\epsilon > 0$ there exists a $\delta = \delta(\epsilon) > 0$ such that if $A \in \Sigma$ and $\nu(A) < \delta$ then

$$\left| \int_A h_n(s)\,\mu(ds) \right| < \epsilon, \qquad n = 1, 2, \ldots.$$

By Egoroff's theorem (8, p. 88) there exists a set $A \in \Sigma$ with $\nu(A) < \delta$ such that $\{h_n(s)\}$ converges to zero uniformly for $s \in S - A$. Having specified ϵ and chosen $\delta = \delta(\epsilon)$ as above, there exists an $N = N(\epsilon)$ such that if $n > N$, then $|h_n(s)| < \epsilon$ for $s \in S - A$. Hence if $n > N$,

$$\left| \int_E h_n(s)\,\mu(ds) \right| \leq \left| \int_{E-A} h_n(s)\,\mu(ds) \right| + \left| \int_{E\cap A} h_n(s)\,\mu(ds) \right| \leq \epsilon\,||\mu||(S) + \epsilon.$$

uniformly for $E \in \Sigma$. Thus the integral is well defined.

Statements (b) and (c) follow readily while (d) and (e) follow from their validity for simple functions, the Vitali-Hahn-Saks theorem, and Corollary 2.4. The first assertion in (f) follows from Definition 2.1. and the observation that $y^* U\mu = (U^* y^*)\mu$, $y^* \in \mathfrak{Y}^*$. The second assertion follows from its validity for simple functions.

2.7. THEOREM. *Let $\{f_n\}$ be a sequence of μ-integrable functions which converges μ-almost everywhere to f. Then f is μ-integrable if*

$$\lim_{||\mu||(E)\to 0} \int_E f_n(s)\,\mu(ds) = 0$$

uniformly for $n = 1, 2, \ldots$. In this case we have

$$\int_E f(s)\,\mu(ds) = \lim_{n\to\infty} \int_E f_n(s)\,\mu(ds).$$

[2] That this theorem remains valid for vector measures may be seen by examining the proof in Saks (15), and has been proved directly by Alexiewicz (1, p. 19) from more general considerations.

Proof. Let k be a fixed positive integer and let $\delta_k > 0$ be such that if $E \in \Sigma$ and $||\mu||(E) < \delta_k$, then

$$\left| \int_E f_n(s)\, \mu(ds) \right| < 2^{-k}, \qquad n = 1, 2, \ldots. \tag{1}$$

This implies that if $||\mu||(E) < \delta_k$, then

$$\int_E |f_n(s)|\, |x^* \mu|(ds) < 16 \cdot 2^{-k}, \quad |x^*| \leqslant 1, \qquad n = 1, 2, \ldots. \tag{2}$$

Evidently, we may assume that $\delta_k \leqslant 2^{-k}$. Let $\eta_k > 0$ be such that if $A \in \Sigma$ and $\nu(A) < \eta_k$ then $||\mu||(A) < \delta_k$. By Corollary 2.4, the sequence $\{f_n\}$ converges ν-almost everywhere, and by an application of Egoroff's theorem, we can select a set $A \in \Sigma$ such that $\nu(A) < \eta_k$ and such that the convergence of $\{f_n\}$ is uniform on $S - A$. Thus if $E \in \Sigma$ and $n, m \geqslant N_k$, we have

$$\begin{aligned} \left| \int_E \{f_n(s) - f_m(s)\}\, \mu(ds) \right| &\leqslant \left| \int_{E-A} \{f_n(s) - f_m(s)\}\, \mu(ds) \right| \\ &+ \left| \int_{E \cap A} f_n(s)\, \mu(ds) \right| \\ &+ \left| \int_{E \cap A} f_m(s)\, \mu(ds) \right| < 2^{-k}\{||\mu||(S)+2\}. \end{aligned} \tag{3}$$

But since k is an arbitrary positive integer, this proves that the sequence $\{\int_E f_n(s)\, \mu(ds)\}$ converges in the norm of $\mathfrak{X}$ for any $E \in \Sigma$.

We now prove that f is μ-integrable. Let δ_k and η_k have the same meaning as in the previous paragraph. Since each f_k is μ-integrable it follows from Egoroff's theorem that there is a simple function g_k and a set $A_k \in \Sigma$ with $\nu(A_k) < \eta_k$ such that

$$|f_k(s) - g_k(s)| < 2^{-k}, \qquad s \in S - A_k, \tag{4}$$

and

$$|g_k(s)| < 2|f_k(s)|, \qquad s \in S. \tag{5}$$

Let

$$B_k = \bigcup_{i=k}^{\infty} A_i$$

so that $B_k \in \Sigma$ and $\{B_k\}$ decreases to the set

$$B = \bigcap_{k=1}^{\infty} B_k.$$

Since

$$||\mu||(B_k) \leqslant \sum_{i=k}^{\infty} ||\mu||(A_i) < \sum_{i=k}^{\infty} \delta_i \leqslant \sum_{i=k}^{\infty} 2^{-i} = 2^{-(k-1)},$$

it follows that $||\mu||(B) = 0$. Now

$$|f(s) - g_k(s)| \leqslant |f(s) - f_k(s)| + f_k(s) - g_k(s)|.$$

If $s \in S - B$, then $s \in S - B_k$ for $k > K(s)$ and so (4) holds provided that $k > K(s)$. Since $\{f_k\}$ was assumed to converge μ-almost everywhere to f, we

conclude that the sequence $\{g_k\}$ converges μ-almost everywhere to f. It remains to show that the integrals $\{\int_E g_n(s)\,\mu(ds)\}$ converge for $E \in \Sigma$. But

$$\left|\int_E \{f_k(s) - g_k(s)\}\,\mu(ds)\right| \leq \left|\int_{E-A_k} \{f_k(s) - g_k(s)\}\,\mu(ds)\right| + \left|\int_{E\cap A_k} f_k(s)\mu(ds)\right| + \left|\int_{E\cap A_k} g_k(s)\,\mu(ds)\right|.$$

The integral over $E - A_k$ is at most $2^{-k}||\mu||(S)$ by (4). Since $||\mu||(E \cap A_k) < \delta_k$, we have seen in (1) that the second term is at most 2^{-k}. In addition, from (5) and (2),

$$\begin{aligned}\left|\int_{E\cap A_k} g_k(s)\,\mu(ds)\right| &= \sup_{|x^*|\leq 1}\left|\int_{E\cap A_k} g_k(s)\,x^*\mu(ds)\right| \\ &\leq \sup_{|x^*|\leq 1}\int_{E\cap A_k} |g_k(s)|\,|x^*\mu|(ds) \\ &\leq 2\sup_{|x^*|\leq 1}\int_{E\cap A_k} |f_k(s)|\,|x^*\mu|(ds) < 32\cdot 2^{-k}.\end{aligned}$$

Combining these, we conclude that

$$\left|\int_E \{f_k(s) - g_k(s)\}\,\mu(ds)\right| < 2^{-k}\{||\mu||(S) + 33\}. \tag{6}$$

Now if k is given, then combining (3) and (6) we see that if $n, m > \max(k, N_k)$ and $E \in \Sigma$, we have

$$\begin{aligned}&\left|\int_E \{g_n(s) - g_m(s)\}\,\mu(ds)\right| \\ &\qquad < 2\cdot 2^{-k}\{||\mu||(S) + 33\} + 2^{-k}\{||\mu||(S) + 2\} \\ &\qquad < M\cdot 2^{-k}.\end{aligned}$$

Hence the sequence of integrals of the simple functions g_n converges for $E \in \Sigma$, so that f is μ-integrable.

To prove the last statement, let $h_n(s) = f(s) - f_n(s)$, $n = 1, 2, \ldots$; then each h_n is μ-integrable and, by Theorem 2.6(e), equation (*) in the proof of that theorem is valid. The argument there applies in this case to prove that

$$\lim_{n\to\infty}\left|\int_E \{f(s) - f_n(s)\}\,\mu(ds)\right| = 0,$$

which is the desired conclusion.

We now show that the theorem on dominated convergence is valid for vector measures.

2.8. THEOREM. *If $\{f_n\}$ is a sequence of μ-integrable functions which converges μ-almost everywhere to f and if g is a μ-integrable function such that $|f_n(s)| \leq g(s)$ μ-almost everywhere, $n = 1, 2, \ldots$, then f is μ-integrable and*

$$\int_E f(s)\,\mu(ds) = \lim_{n\to\infty}\int_E f_n(s)\,\mu(ds), \qquad E \in \Sigma.$$

Proof. By the preceding theorem, we have only to show that

$$\lim_{||\mu||(E)\to 0} \int_E f_n(s)\,\mu(ds) = 0$$

uniformly for $n = 1,2,\ldots$. Now by Theorem 2.6(e), given $\epsilon > 0$, choose $\delta > 0$ such that if $E \in \Sigma$, $||\mu||(E) < \delta$ then

$$\left| \int_E g(s)\,\mu(ds) \right| < \epsilon.$$

Hence if $||\mu||(E) < \delta$ we have

$$\int_E g(s)\,|x^*\mu|(ds) < 4\epsilon, \qquad |x^*| \leqslant 1.$$

Consequently, if $n = 1, 2, \ldots,$ and $||\mu||(E) < \delta$, then

$$\left| \int_E f_n(s)\,\mu(ds) \right| \leqslant \sup_{|x^*| \leqslant 1} \int_E |f_n(s)|\,|x^*\mu|(ds)$$

$$\leqslant \sup_{|x^*| \leqslant 1} \int_E g(s)|x^*\mu|(ds) \leqslant 4\epsilon,$$

from which the conclusion follows.

While the following theorem will not be used in the sequel, it has some independent interest.

2.9. THEOREM. *The range of a vector measure $\mu : \Sigma \to \mathfrak{X}$ is a conditionally weakly compact subset of $\mathfrak{X}$.*

Proof. We shall prove that if $\mathfrak{R} = \{\mu(E) \mid E \in \Sigma\}$ is regarded as a subset of $\mathfrak{X}^{**}$ in the natural embedding, then $\mathfrak{R}$ is conditionally compact in the weak topology of $\mathfrak{X}^{**}$. Since the embedding of $\mathfrak{X}$ in $\mathfrak{X}^{**}$ is closed in this topology, the statement will follow. Now we have seen in Lemma 2.3 that the mapping $U : \mathfrak{X}^* \to \mathrm{ca}(\Sigma)$ defined by $Ux^* = x^*\mu$ is a weakly compact operator; hence the adjoint operator $U^* : \mathrm{ca}^*(\Sigma) \to \mathfrak{X}^{**}$ is also weakly compact. But the unit sphere of $\mathrm{ca}^*(\Sigma)$ certainly contains the linear functionals $\{\phi_E \mid E \in \Sigma\}$ defined by $\phi_E(\lambda) = \lambda(E)$, $\lambda \in \mathrm{ca}(\Sigma)$. Clearly $U^*\{\phi_E \mid E \in \Sigma\} = \{\mu(E) \mid E \in \Sigma\}$, this latter set being regarded in $\mathfrak{X}^{**}$, and $\mathfrak{R}$ is therefore conditionally weakly compact.

3. Representation of operators on continuous functions. Throughout this section S will denote a compact Hausdorff space and Σ its *Borel field*, i.e., the σ-field generated by the closed sets of S. The σ-field generated by the closed G_δ sets of S is called the *Baire field* of S and is denoted by Σ_0. The Banach space of all continuous scalar valued functions on S with the supremum norm is written $C(S)$. The Riesz representation theorem (see Kakutani **(10)**) asserts that the conjugate space of $C(S)$ is isometrically isomorphic to either $\mathrm{ca}(\Sigma_0)$ or to the space $R(S)$ of regular Borel measures on S, where the norm in these measure spaces is the total variation over S.

Let T be a bounded linear operator mapping $C(S)$ into a Banach space $\mathfrak{X}$. We say that T is *compact*, or that it is *weakly compact*, if T maps bounded sets of $C(S)$ into subsets of $\mathfrak{X}$ which are conditionally compact, or which are conditionally weakly compact, respectively. We shall make frequent use of the fact that T is compact (or weakly compact) if and only if its adjoint T^* is compact (or weakly compact). In addition, T is weakly compact if and only if its second adjoint T^{**} maps $C^{**}(S)$ into $\mathfrak{X}$ (more precisely, if the range of T^{**} is contained in the natural embedding of $\mathfrak{X}$ in $\mathfrak{X}^{**}$). For the proofs of the above facts, see **(2; 4)**.

Motivated by the Riesz theorem we are led to inquire when an operator $T : C(S) \to \mathfrak{X}$ may be represented by an integral with respect to a vector measure—we shall see that this is possible if and only if T is weakly compact. First, however, it will be convenient to give a representation of the general operator.

3.1. THEOREM. *If T is an operator on $C(S)$ to $\mathfrak{X}$, there exists a unique set function $\mu : \Sigma \to \mathfrak{X}^{**}$ such that*

(a) $\mu(\cdot)x^* \in R(S)$ *for each* $x^* \in \mathfrak{X}^*$;

(b) *the mapping $x^* \to \mu(\cdot)x^*$ of $\mathfrak{X}^*$ into $R(S)$ is continuous with the $\mathfrak{X}$ and $C(S)$ topologies in these spaces, respectively;*

(c) $x^*Tf = \int_S f(s)\,\mu(ds)x^*$, $f \in C(S)$, $x^* \in \mathfrak{X}^*$;

(d) $|T| = \|\mu\|(S)$, *the semi-variation of μ over S.*

Conversely, if μ is a set function on Σ to $\mathfrak{X}^{**}$ satisfying (a) and (b), then equation (c) defines an operator $T : C(S) \to \mathfrak{X}$ with norm given by (d), and such that $T^*x^* = \mu(\cdot)x^*$.

Proof. For E in Σ let $\mu(E) = T^{**}(\phi_E)$ where ϕ_E is that element of $C^{**}(S)$ defined by the equation $\phi_E(\lambda) = \lambda(E)$ for λ in $R(S)$. Now the proof proceeds along standard lines and we omit the details.

We now ask when we can be assured that the values of μ are contained in $\mathfrak{X}$.

3.2. THEOREM. *If T is a weakly compact operator from $C(S)$ to $\mathfrak{X}$, then there exists a unique vector measure μ on the Borel sets Σ to $\mathfrak{X}$ such that*

(a) $Tf = \int_S f(s)\,\mu(ds)$, $f \in C(S)$;

(b) $|T| = \|\mu\|(S)$;

(c) $T^*x^* = x^*\mu$, $x^* \in \mathfrak{X}^*$.

Conversely, if μ is a vector measure defined on the Baire sets Σ_0 to $\mathfrak{X}$ and if T is defined by (a), then T is a weakly compact operator from $C(S)$ to $\mathfrak{X}$ with norm given by (b) and adjoint operator given by (c).

Proof. If T is weakly compact, then T^{**} maps $C^{**}(S)$ into $\mathfrak{X}$ and thus, from the definition of μ, it is clear that $\mu(E)$ is in $\mathfrak{X}$ for every E in Σ. It follows that $x^*\mu$ is in $R(S)$ and μ is therefore a vector measure. Thus the integral $\int_S f(s)\,\mu(ds)$ exists for f in $C(S)$. From Theorem 3.1 it follows that $x^*T = T^*x^* = x^*\mu$. Thus equation (a) follows from Theorem 2.6(f). Conversely, if $\mu : \Sigma_0 \to \mathfrak{X}$

is a vector measure and if T is defined by (a), then T^* maps x^* into $x^*\mu$. By Lemma 2.3, T^* is weakly compact. Consequently, T is weakly compact.

The next three results were proved by Grothendieck **(7)** who used other methods.[3]

3.3. COROLLARY. *If T is a weakly compact operator from $C(S)$ to $\mathfrak{X}$, then T maps weakly fundamental sequences into strongly convergent sequences. Consequently, T maps conditionally weakly compact subsets of $C(S)$ into conditionally strongly compact subsets of $\mathfrak{X}$.*

Proof. If $\{f_n\}$ is a weakly fundamental sequence in $C(S)$, then $|f_n(s)| < M$ for some number M, and $f_0(s) = \lim f_n(s)$ exists for each $s \in S$, although f_0 may not be in $C(S)$. From Theorem 2.8 we conclude that

$$Tf_n = \int_S f_n(s)\,\mu(ds) \to \int_S f_0(s)\,\mu(ds)$$

in the norm of $\mathfrak{X}$. This proves the first assertion; the second follows directly.

3.4. COROLLARY. *If $U : \mathfrak{Y} \to C(S)$ and $T : C(S) \to \mathfrak{X}$ are weakly compact operators, then $TU : \mathfrak{Y} \to \mathfrak{X}$ is strongly compact.*

3.5. THEOREM.[4] *If T is an arbitrary operator on $C(S)$ to $\mathfrak{X}$, and if $\mathfrak{X}$ is weakly complete,[5] then there exists a vector measure μ_0 defined on the Baire sets Σ_0 of S with values in $\mathfrak{X}$ such that*

$$Tf = \int_S f(s)\,\mu_0(ds), \qquad f \in C(S).$$

Consequently, an arbitrary operator from $C(S)$ to a weakly complete Banach space is weakly compact.

Proof. Let $\mu : \Sigma \to \mathfrak{X}^{**}$ be the set function whose existence was established in Theorem 3.1. We shall show that μ maps Σ_0 into $\mathfrak{X}$. Let f be a bounded function in the first Baire class; the reader may readily verify that there exists a bounded sequence $\{f_n\}$ in $C(S)$ which converges pointwise to f. Then $\{f_n\}$ and $\{Tf_n\}$ are weakly fundamental; hence $\{Tf_n\}$ converges weakly to an element of $\mathfrak{X}$. If f is regarded as an element of $C^{**}(S)$, it is evident that $\{Tf_n\}$ converges to $T^{**}f$ in the $\mathfrak{X}^*$ topology of $\mathfrak{X}^{**}$. We conclude that $T^{**}f \in \mathfrak{X}$. By induction, T^{**} maps the bounded Baire functions into $\mathfrak{X}$; in particular, this is true for the characteristic functions of Baire sets. Hence μ maps Σ_0 into $\mathfrak{X}$. Since the Baire

[3] Although Grothendieck **(7)** did not employ the representation for weakly compact operators given above, he noted the one-to-one correspondence between weakly compact operators on $C(S)$ to $\mathfrak{X}$ and vector measures on the Baire subsets of S to $\mathfrak{X}$. Theorem 3.2 can be proved using results in **(7)** but there is some interest in the measure theoretic approach given here.

[4] Gelfand **(6)** showed that an operator on $C[0, 1]$ to a weakly complete space can be represented as an integral with respect to a vector function of bounded variation. The final conclusion of the present theorem was announced by Pettis **(12)**; it was also proved by Grothendieck.

[5] A Banach space $\mathfrak{X}$ is *weakly complete* if every weakly fundamental sequence in $\mathfrak{X}$ converges weakly to an element of $\mathfrak{X}$.

sets are sufficient to integrate continuous functions, we have the representation

$$Tf = \int_S f(s)\, \mu_0(ds), \qquad f \in C(S),$$

where μ_0 is the restriction of μ to Σ_0. The second assertion follows from the converse part of Theorem 3.2.

3.6. COROLLARY. *An arbitrary continuous linear mapping from an abstract M-space to an abstract L-space is weakly compact.*

Proof. The terminology is that used by Kakutani **(9; 10)**. We observe that the conjugate of an abstract L-space is an abstract M-space with a unit element and hence is isometrically isomorphic to $C(S)$ for some compact Hausdorff space S **(10**, pp. 1023, 998**)**. Further, the adjoint of an M-space is an L-space and is therefore weakly complete **(10**, p. 1021; **9**, p. 537**)**. Thus T^* and T are weakly compact.

3.7. COROLLARY.[6] *If T is a weakly compact map of an abstract M-space (or an abstract L-space) into itself, then T^2 is strongly compact.*

Proof. The second conjugate of an abstract M-space is an abstract M-space with unit **(10**, p. 1023**)** and hence is isometrically isomorphic to $C(S)$. From Corollary 3.4 we conclude that $(T^{**})^2$ and hence T^2 is strongly compact. The second assertion follows from the first and the fact that the conjugate of an L-space is an M-space.

We now turn to the representation of a compact operator on $C(S)$.

3.8. THEOREM. *An operator $T : C(S) \to \mathfrak{X}$ is compact if and only if the vector measure $\mu : \Sigma \to \mathfrak{X}$ corresponding to it (as in Theorem 3.2) takes its values in a compact subset of $\mathfrak{X}$.*

Proof. If T^{**} is compact, then by the construction of μ, we see that the condition is necessary. To see that it is sufficient, it is enough to prove that the set $\mathfrak{K}$ of all sums of the form

$$\sum_{i=1}^{n} \alpha_i\, \mu(E_i),$$

where the $\{E_i\}$ are disjoint and $|\alpha_i| \leqslant 1$, is a totally bounded set in $\mathfrak{X}$. Let $\epsilon > 0$ be given and let M be the semi-variation of μ on S. Select a set $\{\beta_1, \ldots, \beta_p\}$ of complex numbers with $|\beta_i| \leqslant 1$ such that if $|\alpha| \leqslant 1$ then there exists a $\beta_i = \beta(\alpha)$ with $|\beta(\alpha) - \alpha| < \epsilon/2M$. Let $\{F_1, \ldots, F_q\} \subset \Sigma$ be such that if

[6] This result was proved for a concrete L-space by Dunford and Pettis **(3)**, and also by Phillips **(13)**, by explicitly representing the weakly compact operators. It implies that if an abstract M- or L-space contains an infinite dimensional reflexive subspace, then there is no bounded projection mapping onto this subspace.

$E \in \Sigma$, then there is an $F_k = F(E)$ with $|\mu(F(E)) - \mu(E)| < \epsilon/2p$. Then from the definition of the semi-variation

$$\left| \sum_{i=1}^{n} \alpha_i \mu(E_i) - \sum_{i=1}^{n} \beta(\alpha_i) \mu(E_i) \right| = \left| \sum_{i=1}^{n} \{\alpha_i - \beta(\alpha_i)\} \mu(E_i) \right| \leqslant \frac{\epsilon}{2M} \cdot M = \tfrac{1}{2}\epsilon.$$

Further,

$$\sum_{i=1}^{n} \beta(a_i) \mu(E_i)$$

can be written as a sum

$$\sum_{j=1}^{p} \beta_j \mu(E_j'),$$

with $\{E_j'\}$ a disjoint family in Σ. Thus

$$\left| \sum_{j=1}^{p} \beta_j \mu(E_j') - \sum_{j=1}^{p} \beta_j \mu(F(E_j')) \right| < \sum_{j=1}^{p} |\mu(E_j') - \mu(F(E_j'))| < p \cdot \frac{\epsilon}{2p} = \tfrac{1}{2}\epsilon.$$

We have shown that each element in $\mathfrak{R}$ can be approximated within ϵ by sums of the form

$$\sum_{j=1}^{p} \beta_j \mu(F_{k_j}),$$

so that $\mathfrak{R}$ is totally bounded.

4. Special cases. It is easy **(2)** to give representations of operators which map a Banach space into $C(Q)$, where $Q = \{t\}$ is a compact Hausdorff space. Thus an operator $T : C(S) \to C(Q)$ may be studied from two standpoints. We now give kernel representations of operators in these spaces. We remark that in the case that $S = Q = [a, b]$, the general and compact operator was represented by Radon **(14)**, and a representation of the weakly compact operator similar to Theorem 4.2 was given by Sirvint **(16)**. In what follows S and Q are compact Hausdorff spaces.

4.1 THEOREM. *If T is an arbitrary bounded operator from $C(S)$ to $C(Q)$ then T can be represented by the formula*

$$(*) \qquad (Tf)(t) = \int_S f(s)\, K(ds, t), \quad f \in C(S), \quad t \in Q;$$

where K is defined on $\Sigma \times Q$ to the scalar field and satisfies

(i) *$K(\cdot, t) \in R(S)$ for each $t \in Q$;*

(ii) *the integral in (*) is in $C(Q)$ for each $f \in C(S)$;*

(iii) $\sup_{t \in Q} |K(\cdot, t)| = |T| < \infty$.

Conversely, if K satisfies these conditions, the operator defined by (*) maps $C(S)$ into $C(Q)$ and has norm given by (iii).

4.2. THEOREM. *If T is a weakly compact operator from $C(S)$ to $C(Q)$ then T can be represented by formula* (*), *where K satisfies* (i) *and*

(ii′) $K(E, \cdot) \in C(Q)$ *for each* $E \in \Sigma$;

(iii′) $|K(E, t)| \leqslant |T| < \infty$ *for* $E \in \Sigma$, $t \in Q$.

Conversely if K satisfies these conditions, the operator defined by (*) *is a weakly compact map of $C(S)$ into $C(Q)$ and has norm given by* (iii).

Proof. It is easily seen from Theorem 3.2 that a weakly compact operator must have this form. To prove the converse, let $\mu : \Sigma \to C(Q)$ be defined by $\mu(E) = K(E, \cdot)$. Then if $\{E_i\}$ is a disjoint sequence in Σ, it follows from (i) that

$$K\Big(\bigcup_{i=1}^{\infty} E_i, t\Big) = \sum_{i=1}^{\infty} K(E_i, t), \qquad t \in Q,$$

and from the well-known criterion for weak convergence in $C(Q)$ that $x^*\mu$ is countably additive for any $x^* \in C^*(Q)$. With (iii′) this implies that μ is a vector measure in the sense of Definition 2.1.

For some purposes the following representation may be more convenient. It is derived readily from the above and the Radon-Nikodým theorem.

4.3. THEOREM. *If $T : C(S) \to C(Q)$ is weakly compact, then T can be represented by the formula*

$$(+) \qquad (Tf)(t) = \int_S f(s)\, k(s, t)\, \nu(ds), \qquad f \in C(S), \quad t \in Q;$$

where ν is a positive Borel measure on (S, Σ) and k is a function from $S \times Q$ to the scalar field such that

(a) $k(\cdot, t) \in L(S, \Sigma, \nu)$ *for each* $t \in Q$;

(b) $\int_E k(s, \cdot)\, \nu(ds) \in C(Q)$ *for each* $E \in \Sigma$;

(c) $\sup_{t \in Q} \int_S |k(s, t)|\, \nu(ds) = |T| < \infty$.

Conversely, if k and ν satisfy these conditions then (+) *defines a weakly compact operator.*

4.4. THEOREM. *If $T : C(S) \to C(Q)$ is compact, then T can be represented by the formula* (+) *where ν is a positive Borel measure on (S, Σ) and k satisfies* (a), (c) *and*

$$(b') \qquad \lim_{t \to t_0} \int_S |k(s, t) - k(s, t_0)|\, \nu(ds) = 0, \qquad t_0 \in Q.$$

Conversely, if k and ν satisfy these conditions then (+) *defines a compact operator.*

References

1. A. Alexiewicz, *On sequences of operations* (I), Studia Math., *11* (1949), 1–30.
2. R. G. Bartle, *On compactness in functional analysis*, Trans. Amer. Math. Soc., *79* (1955).
3. N. Dunford and B. J. Pettis, *Linear transformations on summable functions*, Trans. Amer. Math. Soc., *47* (1940), 323–392.
4. N. Dunford and J. Schwartz, *Spectral theory*; forthcoming book.
5. W. F. Eberlein, *Weak compactness in Banach spaces* (I), Proc. Nat. Acad. Sci. U.S.A., *33* (1947), 51–53.
6. I. Gelfand, *Abstrakte Funktionen und lineare Operatoren*, Mat. Sbornik (*4*)*46* (1938), 235–284.
7. A. Grothendieck, *Sur les applications linéares faiblement compactes d'espace du type C(K)*, Can. J. Math., *5* (1953), 129–173.
8. P. R. Halmos, *Measure theory* (New York, 1950).

8a. E. Hille, *Functional analysis and semi-groups* (Amer. Math. Soc. Colloquium Publications, vol. 31, 1948).

9. S. Kakutani, *Concrete representation of abstract (L)-spaces and the mean ergodic theorem*, Ann. Math. (2) *42* (1941), 523–537.
10. ———, *Concrete representation of abstract (M)-spaces*, Ann. Math. (2) *42* (1941), 994–1024.
11. B. J. Pettis, *On integration in vector spaces*, Trans. Amer. Math. Soc., *44* (1938), 277–304.
12. ———, *Absolutely continuous functions in vector spaces* (Abstract), Bull. Amer. Math. Soc., *45* (1939), 677.
13. R. S. Phillips, *On linear transformations*, Trans. Amer. Math. Soc., *48* (1940), 516–541.
14. J. Radon, *Über lineare Funktionaltransformationen und Funktionalgleichungen*, Sitzber. Akad. Wiss. Wien, *128* (1919), 1083–1121.
15. S. Saks, *On some functionals*, Trans. Amer. Math. Soc., *35* (1933), 549–556, 965–970.
16. G. Sirvint, *Weak compactness in Banach spaces*, Studia Math., *11* (1949), 71–94.

Yale University

CONFORMAL MAPS WITH LEAST DISTORTION

H. G. HELFENSTEIN

1. Introduction. Our problem is related to the construction of geographical maps as follows. The reason for using conformal geographical maps is that the scale (viz., the ratio of two corresponding line-elements) does not depend on the direction of these line-elements. In an ideal map the scale, being responsible for the preservation of shape, would also be independent of the points where the line-elements begin. Since this is impossible, except for developable surfaces, one tries to construct maps whose scales are "as constant as possible." According to the precise meaning of this expression several "best" maps are possible. Tchebycheff **(1)** studied the case in which the maximum deviation of the scale from a certain constant is minimized. We shall consider here the problem of minimizing the mean quadratic deviation of the scale from a constant. In order to linearize this problem we have to use, as Tchebycheff did also, the logarithm of the scale instead of the scale itself. We prove the existence of a best map in this sense for a simply connected domain on an arbitrary surface. In addition we give some explicit formulae for computing it.

The definition of the mean quadratic error depends on the choice of the parameters on the surface. We use a "normal" system of isothermic coordinates mapping the given surface on the interior of the unit circle of a complex plane. Then our problem reduces to that of the best approximation of a given function by harmonic functions.

2. Notation. Let D be a finite, simply connected domain with more than one boundary point on a surface S which is sufficiently "smooth." Assume that we can find an isothermic system of parameters mapping D on a plane schlicht domain D' which in turn can be transformed conformally on the interior of the unit circle U of a complex z-plane. If $z = x + iy$, then x and y are again isothermic parameters on D which we call a normal system. It is uniquely determined if we let an arbitrary point M of D correspond to the point $z = 0$ and an arbitrary direction through M to the direction of the positive x-axis. The line-element of D becomes in the x, y system:

$$ds = \frac{1}{\gamma(z)}\sqrt{(dx^2 + dy^2)} = \frac{|dz|}{\gamma(z)},$$

where $\gamma(z)$ is a real positive function. Every other conformal map of D can be obtained as a regular analytic function $w = u + iv = f(z)$ defined in U whose derivative must not vanish. In order to fix the image of U with respect to congruent transformations we require that

$$f(0) = 0, \qquad \arg f'(0) = 0. \tag{1}$$

Received August 16, 1954; in revised form January 28, 1955.

We can prescribe also the value of the scale in the fixed point M:

$$|f'(0)| = \frac{\mu}{\gamma(0)}. \tag{2}$$

The distortion in a point x, y now becomes:

$$m = \frac{|dw|}{ds} = |f'(z)|\,\frac{|dz|}{ds} = |f'(z)|\gamma(z). \tag{3}$$

Since $f'(z) \neq 0$ and $\gamma > 0$ we can put

$$\begin{aligned} \Re \log f'(z) &= \log |f'(z)| = \phi(x, y), \\ \log \gamma(z) &= \varphi(z), \end{aligned} \tag{4}$$

whereupon (3) takes the form

$$m = e^{\phi+\varphi}, \tag{5}$$

ϕ being a harmonic function in U. The smoothness of D is now specified as follows: The function $\varphi(x, y)$, which is determined by D, shall be at least twice differentiable with the derivatives bounded in $|z| < 1$.

If m is to be as constant as possible in U then the same must be true for $\phi + \varphi$. We shall therefore determine a harmonic function ϕ and a constant C such that

$$I = \int_U \int (\phi + \varphi - C)^2 dx\, dy \tag{6}$$

becomes a minimum subject to the above mentioned side conditions for $f(z)$. In particular (2) reads as follows:

$$\log|f'(0)| = \phi(0) = \log \frac{\mu}{\gamma(0)}. \tag{7}$$

3. Existence theorem. *If $\varphi(z) = \varphi(R, \theta)$ belongs to the class $C^{(2)}$ in $|z| < 1$,*

(a) *There exists a function $\phi(R, \theta)$ harmonic in $|z| < 1$ with*

$$\phi(0) = \log \frac{\mu}{\gamma(0)}$$

and a constant C which minimize the integral (6).

(b) *If*

$$\phi(R, \theta) - C = \sum_{k=-\infty}^{\infty} b_k(R)\, e^{ik\theta}$$

then $b_k = A_k R^k$,

where

$$A_k = -\frac{k+1}{\pi} \int_0^1 \int_0^{2\pi} \varphi(R, \theta)(Re^{-i\theta})^k R\, dR\, d\theta,$$

and

$$C = \log \frac{\mu}{\gamma(0)} - A_0.$$

(c) *The mapping function subject to conditions* (1) *and* (2) *is given by*

$$f(z) = \frac{\mu}{\gamma(0)} \int_0^z \exp\left(2 \sum_{k=1}^{\infty} A_k t^k\right) dt.$$

Proof. Introduce polar coordinates by

$$x = R \cos\theta, \quad y = R \sin\theta,$$

and expand in Fourier Series

$$\varphi = \sum_{k=-\infty}^{+\infty} a_k(R)\, e^{ik\theta}, \qquad a_k(R) = \frac{1}{2\pi} \int_0^{2\pi} \varphi\, e^{-ik\theta} d\theta, \tag{8}$$

$$k = 0, \pm 1, \pm 2, \ldots,$$

$$\phi + \varphi - C = \sum_{k=-\infty}^{+\infty} C_k(R)\, e^{ik\theta}. \tag{9}$$

Since these functions are real we have

$$a_{-k} = \bar{a}_k, \qquad C_{-k} = \bar{C}_k \tag{10}$$

for every k; since they are one-valued we get

$$a_k(0) = 0, \quad C_k(0) = 0 \tag{11}$$

for $k \neq 0$, and

$$a_0(0) = \varphi(0), \quad C_0(0) = \phi(0) + \varphi(0) - C. \tag{12}$$

Combining (8) and (9) we obtain

$$\phi = \sum_{-\infty}^{+\infty} [C_k(R) - a_k(R)]\, e^{ik\theta} + C, \tag{13}$$

and integrating the condition $\nabla^2 \phi = 0$ under the side conditions (10)–(12), we find, for $k > 0$,

$$C_k(R) = A_k R^k + a_k(R), \quad C_{-k}(R) = \overline{C_k(R)} \tag{14}$$

where the sequence of the constants of integration A_k ($k = 0, 1, 2, \ldots$) is to be determined by our minimum postulate.

We use now the following identity which is true for every pair of complex numbers P and Q:

$$|P|^2 + P\bar{Q} + \bar{P}Q = |P + Q|^2 - |Q|^2.$$

Applying it to

$$P = \frac{A_k}{\sqrt{2(k+1)}}, \quad Q = \sqrt{2(k+1)} \int_0^1 a_k(R)\, R^{k+1} dR,$$

we obtain from (14):

$$\int_0^1 |C_k(R)|^2 R\, dR = \left| \frac{A_k}{\sqrt{2(k+1)}} + \sqrt{2(k+1)} \int_0^1 a_k(R)\, R^{k+1} dR \right|^2 - 2(k+1) \left| \int_0^1 a_k(R)\, R^{k+1} dR \right|^2 + \int_0^1 |a_k(R)^2| R\, dR. \tag{15}$$

The completeness relation yields now for the integral (6):

$$I = 2\pi \int_0^1 |C_0(R)|^2 R\, dR + 4\pi \sum_{k=1}^{\infty} \int_0^1 |C_k(R)|^2 R\, dR, \tag{16}$$

where the interchange of summation and integration will be justified later. Taking into account (15), (16) assumes the form:

$$\begin{aligned} I = {} & 2\pi \left| \frac{A_0}{\sqrt{2}} + \sqrt{2} \int_0^1 a_0(R)\, R\, dR \right|^2 - 4\pi \left| \int_0^1 a_0(R)\, R\, dR \right|^2 \\ & + 2\pi \int_0^1 a_0^2(R)\, R\, dR \\ & + 4\pi \sum_{k=1}^{\infty} \left\{ \left| \frac{A_k}{\sqrt{2(k+1)}} + \sqrt{2(k+1)} \int_0^1 a_k(R)\, R^{k+1} dR \right|^2 \right. \\ & \left. - 2(k+1) \left| \int_0^1 a_k(R)\, R^{k+1} dR \right|^2 + \int_0^1 |a_k(R)|^2 R\, dR \right\}. \end{aligned} \tag{17}$$

Here the unknown constants appear only within squares. Consequently I assumes its minimum for the following values of the constants:

$$A_k = -2(k+1) \int_0^1 a_k(R)\, R^{k+1} dR, \qquad k = 0, 1, 2, \ldots, \tag{18}$$

or, according to (8):

$$A_k = -\frac{k+1}{\pi} \int_0^1 \int_0^{2\pi} \varphi(R, \theta) (R e^{-i\theta})^k R\, dR\, d\theta. \tag{19}$$

Combining (11), (12), and (14) one recognizes that the constant C is also completely determined:

$$C = \phi(0) - A_0 = \log \frac{\mu}{\gamma(0)} + \frac{1}{\pi} \int_0^1 \int_0^{2\pi} \varphi(R, \theta)\, R\, dR\, d\theta.$$

From equations (13) and (18) one concludes that

$$\phi = \log \frac{\mu}{\gamma(0)} + 2 \sum_{k=1}^{\infty} R^k (A_k' \cos k\theta - A_k'' \sin k\theta),$$

where $$A_k' = \Re A_k \quad \text{and} \quad A_k'' = \Im A_k.$$

From this expression one finds the conjugate harmonic function in the usual way, taking into account (1) and (2) in order to determine the constants of integration. This finally leads to

$$f(z) = \frac{\mu}{\gamma(0)} \int_0^z \exp\left(2 \sum_1^{\infty} A_k t^k \right) dt. \tag{20}$$

4. Convergence. The convergence of the power series

$$\sum_1^{\infty} A_k z^k \quad \text{in } |z| < 1 \tag{21}$$

follows from known estimates of the Fourier coefficients. If φ is h–1 times differentiable with respect to θ and possesses a piecewise continuous derivative

of order h with a bound M_h, then h-times repeated integration by parts of (8) yields:

$$|a_k(R)| \leqslant M_h/k^h.$$

Combining this with (18) we have

$$|A_k| < 2\,M_h/k^h. \tag{22}$$

For $|z| \leqslant r_0 < 1$ the series $\sum A_k z^k$ is therefore majorized by $\sum 2M_0 r_0^k$ which is independent of z. The boundness of φ is thus seen to be sufficient for the circle of convergence of (21) to contain U.

If $h = 2$ (as previously assumed) then (21) is also uniformly convergent on the boundary of U (which of course cannot prevent the function $f(z)$ from possibly having a singularity on the boundary).

From (14) and (22) with $h = 1$ one concludes the uniform convergence of

$$\sum_1^\infty |C_k(R)|^2, \qquad 0 \leqslant R \leqslant 1,$$

which has been used in the interchange of summation and integration in (16).

5. Mean value theorem. *For every integer $k > 0$ the mean value of the function $\varphi(R, \theta)R^k e^{\pm ik\theta}$ on the unit circle $|z| < 1$ is numerically equal but opposite in sign to the mean value of $\phi(R, \theta)R^k e^{\pm ik\theta}$.*

Proof. Equation (13) yields:

$$C_k(R) - a_k(R) = \frac{1}{2\pi}\int_0^{2\pi} \phi(R, \theta)\, e^{-ik\theta} d\theta, \qquad k = 0, \pm 1, \pm 2, \ldots$$

hence we have from (14) and (19)

$$A_k R^k = -R^k \cdot \frac{k+1}{\pi}\int_0^1 \int_0^{2\pi} \varphi(r, \theta)\, r^k e^{-ik\theta} r\, dr\, d\theta = \frac{1}{2\pi}\int_0^{2\pi} \phi(R, \theta)\, e^{-ik\theta} d\theta \qquad (k > 0).$$

Multiplying this last equation by R^{k+1} and integrating with respect to R from 0 to 1 we obtain the required result.

6. Determination of the power series of the mapping function. The coefficients of the power series

$$f(z) = \frac{\mu}{\gamma(0)}\sum_{n=1}^\infty C_n z^n \tag{23}$$

can be determined by differentiating twice both (20) and (23).

$$\frac{\gamma(0)}{\mu} f'(z) = \exp\left(2\sum_1^\infty A_k z^k\right) = \sum_1^\infty n C_n z^{n-1}, \tag{24}$$

and

$$2\exp\left(2\sum_1^\infty A_k z^k\right) \cdot \sum_1^\infty k\, A_k z^{k-1} = \sum_1^\infty n(n-1)\, C_n z^{n-2}.$$

Substituting the first factor from (24) we get

$$2\Big(\sum_1^\infty n\,C_n z^{n-1}\Big)\Big(\sum_1^\infty k\,A_k z^{k-1}\Big) = \sum_1^\infty n(n-1)\,C_n z^{n-2}.$$

Equating corresponding coefficients we obtain

$$2\sum A_m C_n = j(j-1)\,C_j,$$

where on the left m and n take the values $1, 2, \ldots j-1$, their sum being always j. Hence

$$C_j = \frac{2}{j(j-1)}\sum_{\nu=1}^{j-1}\nu(j-\nu)\,A_{j-\nu}C_\nu, \qquad j = 2, 3, \ldots.$$

Together with $C_1 = 1$ this recursion formula determines all the C_j's from the A_k's.

7. An integral representation of $\log f'(z)$. Instead of using a series we can express our mapping function also by an integral.

Starting from (20)

$$\log\left[\frac{\gamma(0)}{\mu}f'(z)\right] = 2\sum_{k=1}^\infty A_k z^k,$$

we replace on the right the A_k's by their values (19) and interchange formally summation and integration:

$$\log\left[\frac{\gamma(0)}{\mu}f'(z)\right] = -\frac{2}{\pi}\int_0^1\int_0^{2\pi}\varphi(R,\theta)\sum_{k=1}^\infty (k+1)(zRe^{-i\theta})^k R\,dR\,d\theta.$$

Putting $q = z\,R\,e^{-i\theta}$ the series on the right is obtained by differentiating the geometric series

$$\sum_{k=1}^\infty q^{k+1}$$

with respect to q. Hence

$$\log\left[\frac{\gamma(0)}{\mu}f'(z)\right] = \frac{2}{\pi}\int_0^1\int_0^{2\pi}\varphi(R,\theta)\left[1-\frac{1}{(1-zRe^{-i\theta})^2}\right]R\,dR\,d\theta. \tag{25}$$

For $|z| \leqslant r_0 < 1$ and $R \leqslant 1$ we have $|q| \leqslant r_0 < 1$; consequently the above series is uniformly convergent. For $|z| < 1$ the double integral in (25) is a proper integral of a continuous function; for $|z| = 1$, however, it does not exist in general, not even as an improper integral for in this case

$$|1 - zRe^{-i\theta}| = |z - Re^{i\theta}|.$$

Equation (25) may be transformed in the following way: Let

$$e^{i\theta} = \zeta, \qquad \varphi(R,\theta) = \psi(R,\zeta).$$

Then, since the function $\zeta\,\psi(R,\zeta)$ is defined continuously for $0 \leqslant R \leqslant 1$ and $|\zeta| = 1$, we can form the expression

$$\chi(R,t) = \frac{1}{2\pi i}\oint\frac{\zeta\psi(R,\zeta)}{\zeta - t}\,d\zeta, \tag{26}$$

where the path of integration is the unit circle of the ζ-plane. For $|t| < 1$ this function is analytic in t, and we have

$$\frac{\partial \chi}{\partial t}(R, t) = \frac{1}{2\pi i} \oint \frac{\zeta \psi(R, \zeta)}{(\zeta - t)^2} d\zeta. \tag{27}$$

Transforming the integral in (24) and using (25) and (26) we obtain

$$\log\left[\frac{\gamma(0)}{\mu} f'(z)\right] = 4 \int_0^1 \left[\frac{\partial \chi}{\partial t}(R, 0) - \frac{\partial \chi}{\partial t}(R, zR)\right] R\, dR. \tag{28}$$

One of the integrations is now relegated to the complex integration (26), which may be simpler.

For instance, let us make the further assumption about ψ, that $\zeta\, \psi(R, \zeta)$ can be extended into the interior of the unit circle of the ζ-plane as a regular analytic function of ζ. Then the integration in (26) can be carried out by Cauchy's formula and yields:

$$\chi(R, t) = t\, \psi(R, t).$$

Hence we have from (28):

$$\log\left[\frac{\gamma(0)}{\mu} f'(z)\right] = 4 \int_0^1 \left[\psi(R, 0) - \psi(R, zR) - zR \frac{\partial \psi}{\partial \zeta}(R, zR)\right] R\, dR. \tag{29}$$

It is remarkable that the minimum value of I can be expressed by means of the function (26).

Let

$$\chi(R, t) = \sum_{k=0}^{\infty} b_k(R)\, t^k.$$

Then it is easily seen from (17) that

$$\begin{aligned} I_{\min} = {} & 2\pi \left\{ \int_0^1 |b_1(R)|^2 R\, dR - 2 \left| \int_0^1 b_1(R)\, R\, dR \right|^2 \right\} \\ & + 4\pi \sum_{k=1}^{\infty} \left\{ \int_0^1 |b_{k+1}(R)|^2 R\, dR - 2(k+1) \left| \int_0^1 b_{k+1}(R)\, R^{k+1} dR \right|^2 \right\}. \end{aligned} \tag{30}$$

8. Example. If the function φ does not depend on θ the above-mentioned condition about the extension of $\zeta\psi$ is satisfied. Hence (29) is applicable and yields

$$\log\left[\frac{\gamma(0)}{\mu} f'(z)\right] = 0, \qquad f(z) = \frac{\mu}{\gamma(0)} \cdot z. \tag{31}$$

Let now D be a circular domain on a sphere. Using stereographic projection from one of the endpoints of the diameter perpendicular to the circle's plane, the line-element, and therefore also the function φ, obviously do not depend on θ. Consequently equation (31) holds, which means (assuming the scale to be 1 at the centre): *For a circular domain on a sphere the conformal mapping with least distortion is its symmetrical stereographic projection.*

9. Unsolved problems. In conclusion we mention three natural questions which we leave unanswered.

I. We have not studied the problem of the best choice of the point M on the surface ("the centre of the map") whose image is $z = 0$. Another choice M_1 instead of M would lead to another normal isothermic system $z_1 = x_1 + iy_1$, mapping D again on the unit circle. Consequently

$$z_1 = L(z) = e^{i\tau}\frac{z - \alpha}{\bar{\alpha}z - 1}, \qquad 0 \leqslant \tau < 2\pi, \qquad |\alpha| < 1,$$

where $L(z)$ is a linear function mapping the unit circle onto itself. M_1 has in the old system the coordinates $z = L^{-1}(0) = \alpha$. For the line-element of D we have

$$ds = \frac{|dz|}{\gamma(z)} = \frac{|dz_1|}{\gamma_1(z_1)} = \frac{|L'(z)|\,|dz|}{\gamma_1[L(z)]},$$

and therefore

$$\gamma_1(z_1) = |L'(z)|\gamma(z), \quad \varphi_1(z_1) = \log \gamma_1(z_1) = \log |L'(z)| + \varphi(z).$$

Substituting this in $I_{\min}(30)$ we obtain a function of the complex quantity α, which we have to minimize by a suitable choice of α. The existence of such a minimum (under the side condition $|\alpha| < 1$) seems difficult to prove in general.

II. For which regions D will our map become a schlicht domain? A general answer to this question is difficult; we can hope that for not too large regions which are not too much curved our map will be schlicht, and perhaps more precise sufficient conditions can be found.

III. *Conformal mapping in the large:* Given two conformally equivalent closed surfaces find the mapping with least distortion (in some sense). Since the family of all the conformal mappings in this case depends on a finite number of parameters our problem reduces to a minimum problem instead of a variational problem. But even so the existence of a best map is not obvious.

Reference

1. Tchebycheff, *Sur la construction des cartes géographiques*; Oeuvres, tome I, pp. 233–236, 239–247.

University of Alberta

AN APPLICATION OF SOME SPACES OF LORENTZ

P. G. ROONEY

1. Introduction. The spaces $\Lambda(\alpha)$ and $M(\alpha)$ were defined by Lorentz **(2)** as follows. Let $0 < \alpha < 1$, $0 < l \leqslant \infty$; let ϕ be measurable on $(0, l)$, and, in case $l = \infty$, let the set where $|\phi(x)| > \epsilon$ have finite measure for each positive ϵ. Define

$$\text{I} \qquad ||\phi(\cdot)||_{\Lambda(\alpha)} = \alpha \int_0^l x^{\alpha-1}\phi^*(x)\,dx$$

where $\phi^*(x)$ is the equi-measurable rearrangement of $|\phi|$ in decreasing order, and

$$\text{II} \qquad ||\phi(\cdot)||_{M(\alpha)} = \sup_E (m(E))^{-\alpha} \int_E |\phi(x)|\,dx, \quad E \subseteq (0, l).$$

The spaces $\Lambda(\alpha)$ and $M(\alpha)$ consist of those ϕ for which

$$||\phi(\cdot)||_{\Lambda(\alpha)} < \infty, \quad ||\phi(\cdot)||_{M(\alpha)} < \infty$$

respectively.

Lorentz **(2; §5)** found, among other things, necessary and sufficient conditions that a given sequence be the moment sequence of a function in either $\Lambda(\alpha)$ or $M(\alpha)$, for $l = 1$. It is the object of this paper to find necessary and sufficient conditions that a function $f(s)$ on $s > 0$ be the Laplace transform of a function in $\Lambda(\alpha)$ or $M(\alpha)$ for $l = \infty$. To this end we make use of the Widder-Post inversion operator,

$$\text{III} \qquad L_{k,t}[f(s)] = \frac{(-1)^k}{k!}\left(\frac{k}{t}\right)^{k+1} f^{(k)}\left(\frac{k}{t}\right),$$

whose theory may be found in **(4; chap. VII)**.

Section 2 of this paper contains the theory for the spaces $\Lambda(\alpha)$, and §3 the theory for the spaces $M(\alpha)$.

Henceforth when $l < \infty$, we shall denote the spaces $\Lambda(\alpha)$, $M(\alpha)$, L_p, over $(0, l)$ by $\Lambda(\alpha, l)$, and their respective norms by $||\phi(\cdot)||_{\Lambda(\alpha,l)}$. We shall continue to denote the spaces $\Lambda(\alpha)$ on $(0, \infty)$ by $\Lambda(\alpha)$ and the norms by $||\phi(\cdot)||_{\Lambda(\alpha)}$.

2. The space $\Lambda(\alpha)$. The first theorem yields some properties of the Laplace transform of a function in $\Lambda(\alpha)$, while the second theorem is the representation theorem.

THEOREM 1. *If $\phi \in \Lambda(\alpha)$, and*

$$f(s) = \int_0^\infty e^{-st}\phi(t)\,dt,$$

Received September 21, 1954. This work was done at the Summer Research Institute of the Canadian Mathematical Congress.

then

$$\int_0^\infty s^{-\alpha}|f(s)|\,ds < \infty.$$

If ϕ is positive and decreasing, then the above condition is necessary and sufficient that $\phi \in \Lambda(\alpha)$.

Proof. Suppose $\phi \in \Lambda(\alpha)$. Then

$$\begin{aligned}\int_0^\infty s^{-\alpha}|f(s)|\,ds &< \int_0^\infty s^{-\alpha}ds \int_0^\infty e^{-st}|\phi(t)|\,dt \\ &= \int_0^\infty |\phi(t)|dt \int_0^\infty e^{-st}s^{-\alpha}ds = \Gamma(1-\alpha)\int_0^\infty t^{\alpha-1}|\phi(t)|\,dt \\ &< \alpha^{-1}\Gamma(1-\alpha)||\phi(\cdot)||_{\Lambda(\alpha)} < \infty.\end{aligned}$$

Conversely, suppose ϕ is positive and decreasing. Then

$$\int_0^\infty s^{-\alpha}|f(s)|\,ds = \alpha^{-1}\Gamma(1-\alpha)\,||\phi(\cdot)||_{\Lambda(\alpha)},$$

and $\phi \in \Lambda(\alpha)$.

THEOREM 2. *Necessary and sufficient conditions that a function $f(s)$, defined for $s > 0$, be the Laplace transform of a function in $\Lambda(\alpha)$ are that*

(1) *f has derivatives of all orders in $(0, \infty)$ and $f^{(k)}(s) \to 0$ as $s \to \infty$ $(k = 0, 1, 2, \ldots)$,*

(2) *$||L_{k,\cdot}[f(s)]||_{\Lambda(\alpha)} < N$, where N is independent of k $(k = 0, 1, 2, \ldots)$.*

Proof of necessity. Let

$$f(s) = \int_0^\infty e^{-st}\phi(t)\,dt, \quad \phi \in \Lambda(\alpha).$$

The necessity of (1) is well known; see **(4; chap. 2, §5)**.

Now by **(4; chap. 7, §6)**,

$$L_{k,t}[f(s)] = \int_0^\infty K(t,u)\,\phi(u)\,du,$$

where $K(t,u) = (k/t)^{k+1}\,e^{-ku/t}(u^k/k!)$. Thus $K(t,u) \geqslant 0$, and

$$\int_0^\infty K(t,u)\,du = \int_0^\infty K(t,u)\,dt = 1.$$

Hence, by[1] **(3; Theorem 3.8.1)**, for each $a > 0$,

$$\int_0^a L_{k,t}[f(s)]^*dt < \int_0^a \phi^*(t)\,dt,$$

and thus by **(3; Theorem 3.4.3)**, for any $a > 0$,

$$\alpha\int_0^a t^{\alpha-1}L_{k,t}[f(s)]^*dt < \alpha\int_0^a t^{\alpha-1}\phi^*(t)\,dt.$$

[1]This theorem, like all of Lorentz's, is stated for the case $l = 1$. However, all of Lorentz's theorems used here with one exception (to be noted later) are true for l infinite, as a glance at the proof shows.

Letting $a \to \infty$, we have

$$||L_{k,\cdot}[f(s)]||_{\Lambda(\alpha)} \leqslant ||\phi(\cdot)||_{\Lambda(\alpha)},$$

and (2) is necessary.

Proof of sufficiency. By **(2; 3.5(7))**, if $g(t)$ is positive and non-increasing

$$\int_0^\infty t^{p-1}|g(t)|^p dt \leqslant K_p\left\{\int_0^\infty |g(t)|\,dt\right\}^p, \quad p > 1.$$

Let $p = 1/\alpha$, $g(t) = t^{\alpha-1} L_{k,t}[f(s)]^*$. Then, the above result yields

$$\int_0^\infty |L_{k,t}[f(s)]|^{1/\alpha} dt = \int_0^\infty \{L_{k,t}[f(s)]^*\}^{1/\alpha} dt$$
$$\leqslant K_{1/\alpha}\left\{\int_0^\infty t^{\alpha-1} L_{k,t}[f(s)]^* dt\right\}^{1/\alpha} \leqslant K_{1/\alpha} N^{1/\alpha}.$$

Hence,

$$||L_{k,\cdot}[f(s)]||_{L(1/\alpha)} \leqslant N'$$

where $N' = K^{\alpha}{}_{1/\alpha}\, N$.

Thus, by **(4; chap. 1, §17, and chap. 7, §15)**, $\phi \in L(1/\alpha)$, and an increasing unbounded sequence $\{k_i\}$ exist such that

(i) $||\phi(\cdot)||_{L(1/\alpha)} \leqslant N'$,

(ii) $f(s) = \int_0^\infty e^{-st}\phi(t)\,dt$

(iii) for any $\psi \in L((1-\alpha)^{-1})$,

$$\lim_{i\to\infty} \int_0^\infty \psi(t)\, L_{k_i,t}[f(s)]\,dt = \int_0^\infty \psi(t)\,\phi(t)\,dt.$$

It remains to be shown that $\phi \in \Lambda(\alpha)$.

But by **(3; Theorem 3.6.1)**, for any $\psi \in \mathrm{M}(\alpha)$,

$$\left|\int_0^\infty \psi(t)\, L_{k,t}[f(s)]\,dt\right| \leqslant ||\psi(\cdot)||_{\mathrm{M}(\alpha)} ||L_{k,t}[f(s)]||_{\Lambda(\alpha)} \leqslant N\,||\psi(\cdot)||_{\mathrm{M}(\alpha)}.$$

Hence, by (iii), and since, by **(2; 1.3(4))**, $L((1-\alpha)^{-1}) \subseteq \mathrm{M}(\alpha)$, for any $\psi \in L((1-\alpha)^{-1})$,

$$\left|\int_0^\infty \psi(t)\,\phi(t)\,dt\right| = \lim_{i\to\infty}\left|\int_0^\infty \psi(t)\, L_{k_i,t}[f(s)]\,dt\right| \leqslant N\,||\psi||_{\mathrm{M}(\alpha)}.$$

Changing ψ to ψ sgn ϕ, we have for any positive $\psi \in L((1-\alpha)^{-1})$

$$\int_0^\infty \psi(t)|\phi(t)|\,dt \leqslant N\,||\psi(\cdot)||_{\mathrm{M}(\alpha)},$$

and thus, by **(3; Theorem 3.4.2)**, for any positive $\psi \in L((1-\alpha)^{-1})$,

$$\int_0^\infty \psi(t)\,\phi^*(t)\,dt \leqslant N\,||\psi(\cdot)||_{\mathrm{M}(\alpha)}.$$

Let $\psi(t) = \alpha t^{\alpha-1}$, $0 < \delta \leqslant t \leqslant R$, $\psi(t) = 0$ otherwise. Then $\psi \in L((1-\alpha)^{-1})$, and $||\psi(\cdot)||_{M(\alpha)} \leqslant 1$. Hence

$$\alpha \int_\delta^R t^{\alpha-1}\phi^*(t)\,dt \leqslant N,$$

and so, letting $\delta \to 0$, $R \to \infty$, we have

$$||\phi(\cdot)||_{\Lambda(\alpha)} < \infty,$$

and $\phi \in \Lambda(\alpha)$.

3. The space $M(\alpha)$. The first theorem of this section yields some properties of the Laplace transform of a function in $M(\alpha)$, while the second theorem is the representation theorem.

THEOREM 3. *If $\phi \in M(\alpha)$, and*

$$f(s) = \int_0^\infty e^{-st}\phi(t)\,dt,$$

then $s^\alpha f(s)$ is bounded for $s > 0$. If ϕ is positive and decreasing, then the condition that $s^\alpha f(s)$ be bounded is necessary and sufficient for $\phi \in M(\alpha)$.

Proof. Let $\phi \in M(\alpha)$. Then if $s > 0$, by (3; Theorem 3.6.1),

$$|f(s)| \leqslant \int_0^\infty e^{-st}|\phi(t)|\,dt \leqslant ||e^{-st}||_{\Lambda(\alpha)}||\phi(\cdot)||_{M(\alpha)}$$
$$= \alpha \int_0^\infty t^{\alpha-1}e^{-st}dt||\phi(\cdot)||_{M(\alpha)} = s^{-\alpha}\Gamma(\alpha+1)||\phi(\cdot)||_{M(\alpha)},$$

and $s^\alpha f(s)$ is bounded.

Conversely, suppose ϕ is positive and decreasing, and $s^\alpha f(s)$ is bounded. Let $\delta > 0$, and

$$\frac{1}{2s} < \delta < \frac{1}{s}.$$

Then

$$\int_0^\delta \phi(t)\,dt \leqslant e^{s\delta}\int_0^\delta e^{-st}\phi(t)\,dt \leqslant e\int_0^\infty e^{-st}\phi(t)\,dt \leqslant Ms^{-\alpha} \leqslant M'\delta^{-\alpha},$$

so that $||\phi(\cdot)||_{M(\alpha)} \leqslant M'$ and $\phi \in M(\alpha)$.

THEOREM 4. *Necessary and sufficient conditions that a function $f(s)$, defined for $s > 0$, be the Laplace transform of a function in $M(\alpha)$ are that*

(1) *f has derivatives of all orders in $(0, \infty)$, $f^{(k)}(s) \to 0$ as $s \to \infty$ $(k = 0, 1, 2, \ldots)$,*

(2) *$||L_{k,\cdot}[f(s)]||_{M(\alpha)} \leqslant N$ where N is independent of k $(k = 0, 1, 2, \ldots)$.*

Proof of necessity. Let

$$f(s) = \int_0^\infty e^{-st}\phi(t)\,dt, \qquad \phi \in M(\alpha).$$

The necessity of (1) is well known.

Now as in Theorem 2,

$$L_{k,l}[f(s)] = \frac{1}{k!}\left(\frac{k}{l}\right)^{k+1}\int_0^\infty e^{-ku/l}u^k\phi(u)\,du = \frac{k^{k+1}}{k!}\int_0^\infty e^{-ku}u^k\phi(tu)\,du.$$

Hence, if $m(E) = \delta$,

$$\delta^{-\alpha}\int_E |L_{k,l}[f(s)]|\,dt \leqslant \frac{\delta^{-\alpha}k^{k+1}}{k!}\int_0^\infty e^{-ku}u^k du\int_E |\phi(tu)|\,dt$$
$$= \frac{k^{k+1}}{k!}\int_0^\infty e^{-ku}u^{k+\alpha-1}du(u\delta)^{-\alpha}\int_{uE}|\phi(t)|\,dt$$

where $uE = \{t|t = uv, v \in E\}$, so that $m(uE) = um(E)$.
Thus

$$\delta^{-\alpha}\int_E |L_{k,l}[f(s)]|\,dt \leqslant \frac{k^{k+1}}{k!}||\phi(\cdot)||_{M(\alpha)}\int_0^\infty e^{-ku}u^{k+\alpha-1}du$$
$$= ||\phi(\cdot)||_{M(\alpha)}\Gamma(k+\alpha)/k^\alpha\Gamma(k).$$

Hence, since $\Gamma(k+\alpha)/k^\alpha\,\Gamma(k)$ is bounded, we have $||L_{k,\cdot}[f(s)]||_{M(\alpha)} \leqslant N$.

Proof of sufficiency. It is clear that

$$||L_{k,\cdot}[f(s)]||_{M(\alpha,l)} \leqslant ||L_{k,\cdot}[f(s)]||_{M(\alpha)}.$$

Further, by[2] (2; Theorem 4), $M(\alpha, l) \subseteq L((1-\alpha')^{-1}, l)$ and

$$||L_{k,\cdot}[f(s)]||_{L((1-\alpha')^{-1},l)} \leqslant K_l||L_{k,\cdot}[f(s)]||_{M(\alpha,l)},$$

for every α', $0 < \alpha' < \alpha$. Let α' be fixed $0 < \alpha' < \alpha$ and let $\{l_i\}$ be a positive increasing unbounded sequence. Then by (4; chap. 1, Theorem 17a), since

$$||L_{k,\cdot}[f(s)]||_{L((1-\alpha')^{-1},l_1)} \leqslant K_{l_1}N$$

there is a function $\phi_1 \in L((1-\alpha')^{-1}, l)$ and an increasing unbounded sequence $\{k_{i1}\}$ such that

$$||\phi(\cdot)||_{L((1-\alpha')^{-1},l_1)} \leqslant K_{l_1}N$$

and

$$\lim_{i\to\infty}\int_0^{l_1}\psi(t)\,L_{k_{i1},1}[f(s)]\,dt = \int_0^{l_1}\psi(t)\,\phi_1(t)\,dt,$$

for every $\psi \in L(1/\alpha', l_1)$. Further, since

$$||L_{k_{i1},\cdot}[f(s)]||_{L((1-\alpha')^{-1},l_2)} \leqslant K_{l_2}N$$

there is, by (4; chap. 1, Theorem 17a), a function $\phi_2 \in L((1-\alpha')^{-1}, l_2)$ and an increasing unbounded sequence $\{k_{i2}\} \subseteq \{k_{i1}\}$ such that

$$||\phi_2(\cdot)||_{L((1-\alpha')^{-1},l_2)} \leqslant K_{l_2}N,$$

and

[2] Lorentz states that this theorem is true for l infinite also. However, this is not the case, as it would imply untrue relations between the L_p spaces.

$$\lim_{t\to\infty}\int_0^{l_2}\psi(t)\,L_{k_{i2},t}[f(s)]\,dt=\int_0^{l_2}\psi(t)\,\phi_2(t)\,dt$$

for every $\psi\in L(1/\alpha',l_2)$. Inductively, since

$$||L_{k_{i\,j-1},\cdot}[f(s)]||_{L((1-\alpha')^{-1},\,l_j)}\leqslant K_{l_j}N$$

there is a function $\phi_j\in L((1-\alpha')^{-1},l_j)$, and an increasing unbounded sequence $\{k_{ij}\}\subseteq\{k_{i\,j-1}\}$ such that

$$||\phi_j(\cdot)||_{L((1-\alpha')^{-1},\,l_j)}\leqslant K_{l_j}N$$

and

$$\lim_{t\to\infty}\int_0^{l_j}\psi(t)\,L_{k_{ij},t}[f(s)]\,dt=\int_0^{l_j}\psi(t)\,\phi(t)\,dt,$$

for every $\psi\in L(1/\alpha',l_j)$.

But, if $j<j'$, $\phi_j(t)=\phi_{j'}(t)$ for almost all t in $0\leqslant t\leqslant l_j$. For $\phi_j-\phi_{j'}\in L((1-\alpha')^{-1},l_j)$, and hence if $\psi\in L(1/\alpha',l_j)$ and $\bar\psi=\psi$, $0\leqslant t\leqslant l_j$, $\bar\psi=0$, $t>l_j$, then since $\bar\psi\in L(1/\alpha',l_{j'})$, and $\{k_{ij'}\}\subseteq\{k_{ij}\}$,

$$\begin{aligned}\int_0^{l_j}\psi(t)(\phi_j(t)-\phi_{j'}(t))\,dt&=\int_0^{l_j}\psi(t)\,\phi_j(t)\,dt-\int_0^{l_{j'}}\bar\psi(t)\,\phi_{j'}(t)\,dt\\&=\lim_{t\to\infty}\int_0^{l_j}\psi(t)\,L_{k_{ij},t}[f(s)]\,dt-\lim_{t\to\infty}\int_0^{l_{j'}}\bar\psi(t)\,L_{k_{ij'},t}[f(s)]\,dt\\&=\lim_{t\to\infty}\left\{\int_0^{l_j}\psi(t)\,L_{k_{ij'},t}[f(s)]\,dt-\int_0^{l_j}\psi(t)\,L_{k_{ij'},t}[f(s)]\,dt\right\}=0.\end{aligned}$$

Thus by (**1**; chap. IV, §4.2 and Theorem 3), $\phi_j(t)=\phi_{j'}(t)$ almost everywhere in $0<t<l_j$.

For each $t>0$ let $\phi(t)=\phi_j(t)$ where j is the least i such that $t<l_i$. Then clearly $\phi\in L((1-\alpha')^{-1},l)$ for each $l>0$, and if $k_i=k_{ii}$, and $\psi\in L(1/\alpha',l)$,

$$\lim_{t\to\infty}\int_0^{l}\psi(t)\,L_{k_i,t}[f(s)]\,dt=\int_0^{l}\psi(t)\,\phi(t)\,dt.$$

Further, ϕ has a Laplace transform. For if $s>0$, then

$$e^{-st}\operatorname{sgn}(\phi(t))\in L(1/\alpha',l)\cap\Lambda(\alpha)$$

and thus by (**3**; Theorem 3.6.1)

$$\begin{aligned}\int_0^{l}e^{-st}|\phi(t)|\,dt&=\left|\int_0^{l}e^{-st}\operatorname{sgn}(\phi(t))\,\phi(t)\,dt\right|\\&=\lim_{t\to\infty}\left|\int_0^{l}e^{-st}\operatorname{sgn}(\phi(t))\,L_{k_i,t}[f(s)]\,dt\right|\\&\leqslant||e^{-st}||_{(\Lambda\alpha)}\limsup_{t\to\infty}||L_{k_i,\cdot}[f(s)]||_{M(\alpha)}\leqslant s^{-\alpha}\Gamma(\alpha+1)\,N.\end{aligned}$$

Thus

$$\int_0^{\infty}e^{-st}\phi(t)\,dt$$

exists for $s > 0$. Also,

$$\lim_{t\to\infty} \int_0^\infty e^{-st} L_{k_t,\,t}[f(s)]\,dt = \int_0^\infty e^{-st}\phi(t)\,dt,$$

for each $s > 0$. For, by (3; Theorem 3.6.1.),

$$\left| \int_l^\infty e^{-st} L_{k_t,\,t}[f(s)]\,dt \right| \leqslant ||L_{k_t,\,t}[f(s)]||_{\mathrm{M}(\alpha)} \cdot \alpha \int_l^\infty t^{\alpha-1} e^{-st}\,dt$$
$$\leqslant N\alpha \int_l^\infty e^{-st} t^{\alpha-1}\,dt < \epsilon$$

and we may also choose l so large that

$$\int_l^\infty e^{-st}|\phi(t)|\,dt < \epsilon.$$

Then,

$$\limsup_{t\to\infty} \left| \int_0^\infty e^{-st}(\phi(t) - L_{k_t,\,t}[f(s)])\,dt \right|$$
$$\leqslant \limsup_{t\to\infty} \left| \int_0^l e^{-st}(\phi(t) - L_{k_t,\,t}[f(s)])\,dt \right| + 2\epsilon = 2\epsilon,$$

and thus since ϵ is arbitrary,

$$\lim_{t\to\infty} \int_0^\infty e^{-st} L_{k_t,\,t}[f(s)]\,dt = \int_0^\infty e^{-st}\phi(t)\,dt.$$

But by (4; chap. 7, Theorem 11b), this last limit is equal to $f(s)$. Thus $f(s)$ is the Laplace transform of ϕ, and all that remains to be shown is that $\phi \in \mathrm{M}(\alpha)$.

But by (4; chap. 7, Theorem 6a)

$$\lim_{k\to\infty} L_{k,\,t}[f(s)] = \phi(t) \text{ a.e.}$$

Hence if E is any subset, of measure δ, then from Fatou's lemma

$$\int_E |\phi(t)|\,dt \leqslant \liminf_k \int_E |L_{k,\,t}[f(s)]|\,dt \leqslant N\delta^\alpha$$

Hence

$$||\phi(t)||_{\mathrm{M}(\alpha)} = \sup_E \delta^{-\alpha} \int_E |\phi(t)|\,dt \leqslant N$$

and $\phi \in \mathrm{M}(\alpha)$.

In conclusion it may be mentioned that results of the type obtained in theorems 2 and 4 hold for considerably more general spaces than $\Lambda(\alpha)$ and $\mathrm{M}(\alpha)$. For example, analogues of these theorems hold true if the values of $f(s)$ be in a reflexive Banach space; the proof of this fact is much like the proofs given here.

References

1. S. Banach, *Theorie des operations lineaires* (Warsaw, 1932).
2. G. G. Lorentz, *Some new functional spaces*, Ann. Math., *51* (1950), 37–55.
3. ———, *Bernstein polynomials* (Toronto, 1953).
4. D. V. Widder, *The Laplace transform* (Princeton, 1941).

University of Alberta

ISOMORPHISMS OF FACTORS OF INFINITE TYPE

R. V. KADISON

1. Introduction. One of the striking results of the work done by Murray and von Neumann **(9)** in the analysis of rings of operators[1] on a Hilbert space is the reduction of the unitary equivalence problem for certain types of factors[2] to the problem of algebraic equivalence. Roughly speaking, they associate with each concrete representation of a factor a number (which measures the relative size of the factor and its commutant)—the so-called "coupling constant." Two factors are unitarily equivalent if and only if they are algebraically isomorphic and have the same coupling constant. Somewhat more precisely, Murray and von Neumann show that the given algebraic isomorphism can be implemented by a unitary transformation. Their results do not apply to factors of type III nor does the result concerning the possibility of implementation of an isomorphism by a unitary transformation apply to the case of II_∞ factors with II_1 commutants. Recently[3] Griffin **(4; 7)** pointed out the surprising fact that (at least in the case of a separable Hilbert space) every isomorphism between factors of type III can be implemented by a unitary transformation. By combining the techniques of Nakano **(10)** and Segal **(11)** in their multiplicity theory of abelian rings of operators with the global ring techniques of Dixmier **(1)** and Kaplansky **(6)**, and the Dye-Radon-Nikodym Theorem **(2)**, Griffin **(5)** was able to extend the concept of "coupling constant" from factors in the separable case to "coupling operator" for rings of operators on an arbitrary Hilbert space. He thereby extended the unitary equivalence results of Murray and von Neumann to rings of operators. However, there does not seem to be a description, in the literature, of the possible isomorphisms between II_∞'s with II_1 commutants (Griffin's results are worded so as to exclude this case). One knows, for example, that each $*$-automorphism (adjoint-preserving automorphism) of a factor is implemented by a unitary transformation of the underlying Hilbert space provided the factor is not of type II_∞ with a II_1 commutant. What the situation is, in this last case, seems to be unknown. This note supplies the missing information concerning isomorphisms between rings of type II_∞ with II_1 commutants. In particular, we show that the group of unitarily induced automorphisms of a factor of type II_∞ with a II_1 commutant is a normal subgroup of the group of $*$-automorphisms

Received September 27, 1954. This research was carried out during the tenure of a Fulbright Grant.

[1] A ring of operators is a weakly closed, self-adjoint algebra of operators on a Hilbert space.

[2] A factor is a ring of operators whose center consists of scalar multiples of its unit element.

[3] The author is indebted to E. L. Griffin for having had this result and proof made available to him in 1952.

and that the quotient group is (canonically) isomorphic to the fundamental group[4] of the II_1 commutant.

The results of the present note will be employed in a forthcoming account of the unitary invariants of representations of arbitrary C*-algebras.

2. The automorphism group. The first question with which we shall deal—the nature of the *-automorphisms of a factor of type II_∞ with a II_1 commutant on a separable Hilbert space—is the simplest one, from a technical viewpoint, but, nevertheless, contains all the essential features of the more general investigation of the next section.

THEOREM 1. *If $\mathfrak{M}$ is a factor of type II_∞ with dimension function D and commutant $\mathfrak{M}'$ of type II_1, then the mapping which takes each *-automorphism ϕ of $\mathfrak{M}$ into $D[\phi(E)]/D(E)$, with E some fixed, finite, non-zero projection in $\mathfrak{M}$ is a group homomorphism of the group, $\mathfrak{G}$, of *-automorphisms of $\mathfrak{M}$ onto the fundamental group of $\mathfrak{M}$ with kernel, $\mathfrak{U}$, consisting of those *-automorphisms of $\mathfrak{M}$ which are implemented by unitary transformations of the underlying Hilbert space, $\mathfrak{H}$.*

Proof. Since $D\cdot\phi$ serves as a dimension function on $\mathfrak{M}$ (for each *-automorphism, ϕ, of $\mathfrak{M}$), $D\cdot\phi$ is a constant multiple, say $\alpha(\phi)$, of D. If E is chosen to be a projection in $\mathfrak{M}$ with $D(E) = 1$, then clearly, $\alpha(\phi) = D[\phi(E)]$. Thus, if η is another *-automorphism of $\mathfrak{M}$, then

$$\alpha(\eta\cdot\phi) = D[\eta\phi(E)] = \alpha(\eta)D[\phi(E)] = \alpha(\eta)\cdot\alpha(\phi);$$

so that α is a group homomorphism of $\mathfrak{G}$ into the group of positive reals. We examine the kernel of α. Suppose then that $\alpha(\phi) = 1$. Since $\mathfrak{M}'$ is of type II_1 and $\mathfrak{M}$ of type II_∞, it is possible (**8**, pp. 178–180) to choose a unit vector x in $\mathfrak{H}$ so that[5] $[\mathfrak{M}x] = \mathfrak{H}$. Let F be the orthogonal projection on the space $[\mathfrak{M}'x]$. Then F lies in $\mathfrak{M}$ and is finite. Moreover $F\mathfrak{M}F$ and $\mathfrak{M}'F$ restricted to the space $[\mathfrak{M}'x]$ are factors of type II_1, one the commutant of the other, with coupling constant 1, since x serves as a cyclic vector for both $F\mathfrak{M}F$ and $\mathfrak{M}'F$ (recall that the total space under consideration, at the moment, is $[\mathfrak{M}'x]$). By assumption on ϕ, however, $D(F) = D[\phi(F)]$, and, since F is finite, one can find a unitary operator, U, in $\mathfrak{M}$ such that $UFU^{-1} = \phi(F)$. Now $[\mathfrak{M}Ux]$ contains $[\mathfrak{M}U^{-1}Ux] = [\mathfrak{M}x] = \mathfrak{H}$, and

$$[\mathfrak{M}'Ux] = [U\mathfrak{M}'x] = U[\mathfrak{M}'x] = U(F(\mathfrak{H})) = \phi(F)(\mathfrak{H}).$$

Thus Ux plays the same role with respect to $\phi(F)$ as x did with respect to F. It follows that $\phi(F)\mathfrak{M}\phi(F)$ and $\mathfrak{M}'\phi(F)$ are factors of type II_1, each the

[4] For the definition of fundamental group of a factor see (**9**). It should be noted that a factor of type II can be viewed as an infinite matrix ring over various factors of type II_1 all belonging to the genus of the II_∞ and, so, all having the same fundamental group which we may call the fundamental group of the given II_∞. This group is also the fundamental group of the commutant.

[5] We denote by $[\mathfrak{M}x]$ the closed subspace spanned by vectors of the form Ax, with A in $\mathfrak{M}$.

commutant of the other, with coupling constant 1. Moreover, ϕ restricted to $F\mathfrak{M}F$ maps this ring isomorphically upon $\phi(F)\mathfrak{M}\phi(F)$. Since $F\mathfrak{M}F$ as represented upon $F(\mathfrak{H})$ and $\phi(F)\mathfrak{M}\phi(F)$ as represented upon $\phi(F)(\mathfrak{H})$ have coupling constant 1, the known theory **(9)** tells us that there is a unitary transformation, U_1, of $F(\mathfrak{H})$ upon $\phi(F)(\mathfrak{H})$ which implements the restricted ϕ. Now choose orthogonal, equivalent projections $F_1, F_2 \ldots$ in $\mathfrak{M}$ with sum I and with $F = F_1$. Let V_i be a partial isometry in $\mathfrak{M}$ with initial space $F_1(\mathfrak{H})$ and final space $F_i(\mathfrak{H})$. (Take $V_1 = F_1$.) The map U_1, discussed above, transforms $F_1(\mathfrak{H})$ onto $\phi(F_1)(\mathfrak{H})$ and implements ϕ restricted to $F_1\mathfrak{M}F_1$. Define U_n to be $\phi(V_n)U_1V_n^*$. Clearly, U_n is a unitary transformation of $F_n(\mathfrak{H})$ onto $\phi(F_n)(\mathfrak{H})$. We assert that U_n implements the isomorphism ϕ restricted to $F_n\mathfrak{M}F_n$. In fact,

$$\begin{aligned} U_nF_nAF_nU_n^{-1} &= \phi(V_n)U_1(V_n^*F_nAF_nV_n)U_1^{-1}\phi(V_n)^* \\ &= \phi(V_n)U_1(F_1V_n^*F_nAF_nV_nF_1)U_1^{-1}\phi(V_n^*) \\ &= \phi(V_n)\phi(F_1V_n^*F_nAF_nV_nF_1)\phi(V_n^*) \\ &= \phi(V_nF_1V_n^*F_nAF_nV_nF_1V_n^*) = \phi(F_nAF_n). \end{aligned}$$

The transformation U defined to be U_n on each of the spaces $F_n(\mathfrak{H})$ is a unitary transformation of $\mathfrak{H}$ onto $\mathfrak{H}$ and certainly implements ϕ on each of the rings $F_n\mathfrak{M}F_n$. Moreover,

$$\begin{aligned} UV_nU^{-1} &= UV_n(\textstyle\sum_k U_k^{-1}) = UV_nU_1^{-1} = (\sum_k U_k)V_nU_1^{-1} = U_nV_nU_1^{-1} \\ &= (\phi(V_n)U_1V_n^*)V_nU_1^{-1} = \phi(V_n)U_1F_1U_1^{-1} = \phi(V_n)\phi(F_1) \\ &= \phi(V_nF_1) = \phi(V_n). \end{aligned}$$

Thus

$$\begin{aligned} UAU^{-1} &= U(\textstyle\sum_n F_n)A(\sum_m F_m)U^{-1} = U(\sum_{n,m} F_nAF_m)U^{-1} \\ &= \textstyle\sum_{n,m} UF_nA\ F_mU^{-1} \\ &= \textstyle\sum_{n,m} UV_nV_n^*AV_mV_m^*U^{-1} \\ &= \textstyle\sum_{n,m} (UV_nU^{-1})(UF_1V_n^*AV_mF_1U^{-1})(UV_m^*U^{-1}) \\ &= \textstyle\sum_{n,m} \phi(V_n)\phi(V_n^*AV_m)\phi(V_m^*) = \sum_n \phi(F_n)\phi(A)\phi(F_m) \\ &= (\textstyle\sum_n \phi(F_n))\phi(A)(\sum_m \phi(F_m)) = \phi(A), \end{aligned}$$

since each $*$-automorphismis countably additive We have established that each automorphism, ϕ, in the kernel of α is induced by a unitary transformation of $\mathfrak{H}$. Suppose, on the other hand, that ϕ is an automorphism of $\mathfrak{M}$ induced by the unitary transformation U of $\mathfrak{H}$. Once again, choosing x a unit vector in $\mathfrak{H}$ such that $[\mathfrak{M}x] = \mathfrak{H}$ and defining F to be the orthogonal projection with range $[\mathfrak{M}'x]$, we have that F lies in $\mathfrak{M}$ and is finite. In addition,

$$[\mathfrak{M}Ux] = [UU^{-1}\,\mathfrak{M}Ux] = U[\mathfrak{M}x] = U(\mathfrak{H}) = \mathfrak{H},$$

and

$$[\mathfrak{M}'\,Ux] = U[U^{-1}\,\mathfrak{M}'\,Ux] = U[\mathfrak{M}'x] = UFU^{-1}\,U(\mathfrak{H}) = \phi(F)(\mathfrak{H}).$$

Now it follows **(8**, pp. 178–180**)** that $[\mathfrak{M}'x]$ is equivalent to $[\mathfrak{M}'\,Ux]$ modulo $\mathfrak{M}$, i.e., $D(F) = D[\phi(F)]$, since $[\mathfrak{M}x] = \mathfrak{H}$ is equivalent to $[\mathfrak{M}\,Ux] = \mathfrak{H}$ modulo $\mathfrak{M}'$, so that $\alpha(\phi) = 1$; and ϕ lies in the kernel of α. Thus we have identified the

kernel of α with the group $\mathfrak{U}$ of unitarily induced automorphisms of $\mathfrak{M}$. Finally we show that the image of α is precisely the fundamental group of $\mathfrak{M}$. In fact, with F a projection of relative dimension 1 in $\mathfrak{M}$, $F\mathfrak{M}F$ is a factor of type II_1, and $\phi(F)\mathfrak{M}\phi(F)$ is its $D[\phi(F)] = \alpha(\phi)$th power[4] (for, if, say $\alpha(\phi) \leqslant 1$, then $\phi(F)\mathfrak{M}\phi(F)$ is unitarily equivalent to the restriction of $F\mathfrak{M}F$ to any projection in $F\mathfrak{M}$ of relative dimension $\alpha(\phi)$). However, ϕ induces an isomorphism of $F\mathfrak{M}F$ upon $\phi(F)\mathfrak{M}\phi(F)$, so that $\alpha(\phi)$ lies in the fundamental group of $F\mathfrak{M}F$ (which is the fundamental group of $\mathfrak{M}$). Suppose that a is in the fundamental group of $\mathfrak{M}$. Choose **(9)** two (infinite, sets of matrix units $[E_{ij}]_{i,j=1,2,\ldots}$ and $[F_{ij}]_{i,j=1,2,\ldots}$ in $\mathfrak{M}$, with E_{nn}, F_{nn} projections of relative dimensions 1 and a, respectively, for each n. If we denote by $\mathfrak{N}_1$ and $\mathfrak{N}_2$ the sets of elements in $\mathfrak{M}$ which commute with $[E_{ij}]$ and $[F_{ij}]$, respectively, then $\mathfrak{N}_1$ and $\mathfrak{N}_2$ are subfactors of $\mathfrak{M}$ of type II_1, and are isomorphic to $E_{11}\mathfrak{M}E_{11}$ and $F_{11}\mathfrak{M}F_{11}$, respectively. Thus $\mathfrak{N}_2$ is the ath power of $\mathfrak{N}_1$, and, since a lies in the fundamental group of $\mathfrak{M}$ (hence, of $\mathfrak{N}_1$), there is an isomorphism η of $\mathfrak{N}_1$ onto $\mathfrak{N}_2$. Now $\mathfrak{M}$ is isomorphic to the denumerably infinite matrix rings over $\mathfrak{N}_1$ and over $\mathfrak{N}_2$ in which only those matrices occur which yield bounded operators on the denumerably infinite direct sum of $\mathfrak{H}$ with itself. Let ϕ_1, ϕ_2 be these isomorphisms and $\mathfrak{N}_1^\circ$, $\mathfrak{N}_2^\circ$ the matrix rings, respectively. If A° is a denumerably infinite matrix over $\mathfrak{N}_1$ (or $\mathfrak{N}_2$) it will act as a bounded operator if and only if the bounds of the operators obtained from A° by replacing the entries whose row or column index exceeds n by 0, forms a bounded set of numbers. Now η extends, in the obvious way, to a $*$-isomorphism η_n of the $n \times n$ matrix ring over $\mathfrak{N}_1$ onto the $n \times n$ matrix ring over $\mathfrak{N}_2$. Then η_n is norm preserving and it follows from the foregoing characterization of the operators in $\mathfrak{N}_1^\circ$, $\mathfrak{N}_2^\circ$ that η°, the extension of η to $\mathfrak{N}_1^\circ$, is a $*$-isomorphism of $\mathfrak{N}_1^\circ$ onto $\mathfrak{N}_2^\circ$. Under ϕ_1 and ϕ_2, respectively, E_{11} and F_{11}, respectively, map onto the matrices in $\mathfrak{N}_1^\circ$ and $\mathfrak{N}_2^\circ$, respectively, whose entry in the first column and row is I and whose other entries are 0. Thus the $*$-automorphism, $\phi_2^{-1}\eta^\circ\phi_1$, of $\mathfrak{M}$ carries E_{11} onto F_{11}. It follows that $\alpha(\phi_2^{-1}\eta^\circ\phi_1) = a$, so that the homomorphism α maps onto the fundamental group of $\mathfrak{M}$.

COROLLARY. *There exist factors of type* II_∞ *with* II_1 *commutants which admit non-unitarily induced automorphisms.*

Proof. Since the fundamental group of the approximately finite II_1 is the multiplicative group of positive reals, the automorphism group of the approximately finite II_∞ (bounded, denumerably infinite matrices over the approximately finite II_1) contains distinct cosets modulo the group of unitarily induced automorphisms corresponding to each positive real number. The denumerably infinite matrix representation, acting in the usual way on the direct sum of Hilbert space with itself a denumerably infinite number of times has a II_1 commutant and provides the desired example.

It is a rather surprising observation that, in a certain sense, the more complicated factors of type II_∞ have the less complicated automorphism groups.

Indeed, one tends to think of a factor of type II_∞ whose fundamental group consists of 1 alone as being quite complicated structurally (the approximately finite factors of type II_∞ would appear to be the least complicated), while, in this case, all $*$-automorphisms are unitarily induced. On the other hand, the only information we have about the fundamental group of any factor is that the fundamental group of an approximately finite factor of type II_1 is the group of positive reals. It may well be that this is the fundamental group of all factors of type II_1. It would be quite interesting to know, for example, the fundamental group of the factor of type II_1 which is the group algebra of the free group on two generators.

3. The linking operator of an isomorphism. In this section, we shall deal with the more general situation of $*$-isomorphisms between rings of type II_∞ with II_1 commutants on arbitrary Hilbert spaces.

DEFINITION. *If ϕ is a $*$-isomorphism between two rings of operators $\mathfrak{M}_1$ and $\mathfrak{M}_2$ of type II_∞ with II_1 commutants acting on the Hilbert spaces $\mathfrak{H}_1$ and $\mathfrak{H}_2$, respectively, we shall call the operator $D[\phi(E)]$ in the center of $\mathfrak{M}_2$, the* linking operator *for ϕ, where E is the projection in $\mathfrak{M}_1$ with range $[\mathfrak{M}_1' x]$, where x is a unit vector such that $[\mathfrak{M}_1 x] = \mathfrak{H}_1$, and where D is the center-valued dimension function on $\mathfrak{M}_2$ normalized so that $D(F) = \mathrm{I}$. (F defined for $\mathfrak{M}_2$ in the same way as E is defined for $\mathfrak{M}_1$; assuming $\mathfrak{M}_1$, $\mathfrak{M}_2$ have countably decomposable centers.*

Several remarks are appropriate with regard to this definition. In the first place, the cyclic vector x exists since $\mathfrak{M}_1$ is of type II_∞ and $\mathfrak{M}_1'$ is of type II_1. Secondly, E is finite with central carrier the identity, and any other projection in $\mathfrak{M}_1$ arising from a cyclic vector such as x is equivalent to E. Since ϕ maps finite projections into finite projections and equivalent projections into equivalent projections, $D[\phi(E)]$ is independent of the choice of x and is a positive operator in the center of $\mathfrak{M}_2$ (observe that $\phi(E)$ is finite and has central carrier I). Using direct sums, we assume our rings have countably decomposable centres.

THEOREM 2. *If ϕ is a $*$-isomorphism of the ring of operators $\mathfrak{M}_1$ of type II_∞ onto the ring of operators $\mathfrak{M}_2$ and the commutants of $\mathfrak{M}_1$ and $\mathfrak{M}_2$ are of type II_1, then ϕ is implemented by a unitary transformation of $\mathfrak{H}_1$ onto $\mathfrak{H}_2$, the Hilbert spaces upon which $\mathfrak{M}_1$ and $\mathfrak{M}_2$, respectively, act, if and only if the linking operator for ϕ is the identity operator.*

Proof. Suppose first that ϕ is implemented by a unitary transformation U of $\mathfrak{H}_1$ onto $\mathfrak{H}_2$. In this case, with the notation of the above definition,

$$[\mathfrak{M}_2\, Ux] = U[U^{-1}\mathfrak{M}_2\, Ux] = U[\mathfrak{M}_1 x] = U(\mathfrak{H}_1) = \mathfrak{H}_2,$$

and

$$[\mathfrak{M}_2'\, Ux] = U[U^{-1}\mathfrak{M}_2'\, Ux] = U[\mathfrak{M}_1' x] = U(E(\mathfrak{H}_1)) = \phi(E)(\mathfrak{H}_2).$$

Thus $D[\phi(E)] = \mathrm{I}$, for $\phi(E)$ arises from a cyclic vector Ux, and is therefore equivalent to the normalizing projection F for D, in $\mathfrak{M}_2$.

We assume now that $D[\phi(E)] = I$, so that $\phi(E)$ is equivalent to the normalizing projection F for D and arises from a cyclic vector for $\mathfrak{M}_2$. Thus $\phi(E)\mathfrak{M}_2\phi(E)$ and $\mathfrak{M}_2'\,\phi(E)$ are of type II_1 with a joint cyclic vector, as are $E\,\mathfrak{M}_1\,E$ and $\mathfrak{M}_1'\,E$. Moreover, ϕ yields a $*$-isomorphism of $E\,\mathfrak{M}_1\,E$ onto $\phi(E)\,\mathfrak{M}_2\,\phi(E)$. This last $*$-isomorphism is implemented **(3; 5)** by a unitary transformation U_1 of $E(\mathfrak{H}_1)$ onto $\phi(E)(\mathfrak{H}_2)$. Again, as in Theorem 1, one can find a (possibly uncountable) family of projections $[E_\alpha]$ in $\mathfrak{M}_1$, mutually orthogonal and equivalent to E, with sum I, and with E as one of the projections of the family. Let V_α be a partial isometry in $\mathfrak{M}_1$ with initial space $E(\mathfrak{H}_1)$ and final space $E_\alpha(\mathfrak{H}_1)$ (let E be the partial isometry with initial and final space $E(\mathfrak{H}_1)$). Define U_α to be $\phi(V_\alpha)U_1\,V_\alpha^*$. As in Theorem 1, the unitary operator U defined as U_α on $E_\alpha(\mathfrak{H}_1)$, for each α, implements ϕ on each of the rings $E_\alpha\,\mathfrak{M}_1\,E_\alpha$ and $UV_\alpha\,U^{-1} = \phi(V_\alpha)$. Thus the isomorphisms $A \to UA\,U^{-1}$ and $A \to \phi(A)$ agree on a subset of $\mathfrak{M}_1$ dense in the strongest topology. Since both mappings are strongestly continuous **(3; 5)**, they agree on $\mathfrak{M}_1$ and ϕ is implemented by the unitary transformation U.

It is now a simple matter to incorporate the above result into the statements of Griffin **(3; 5)** to completely answer the question of when $*$-isomorphisms between arbitrary rings of operators on arbitrary Hilbert spaces are induced by unitary transformations.

References

1. J. Dixmier, *Les anneaux d'opérateurs de classe finie*, Ann. Ecole Norm. (1949), 209–261.
2. H. A. Dye, *The Radon Nikodym theorem for finite rings of operators*, Trans. Amer. Math. Soc., *74* (1952), 243–280.
3. ———, *The unitary structure in finite rings of operators*, Duke Math. J., *40* (1953), 55–70.
4. E. L. Griffin, *Isomorphisms of rings of type III*, Bull. Amer. Math. Soc., *58* (1952), Abs. No. 429.
5. ———, *Some contributions to the theory of rings of operators*, Trans. Amer. Math. Soc., *75* (1953), 471–504.
6. I. Kaplansky, *Projections in Banach algebras*, Ann. Math., *53* (1951), 235–249.
7. Y. Misonou, *Unitary equivalence of factors of type III*, Proc. Jap. Acad., *29* (1952), 482–485.
8. F. J. Murray and J. von Neumann, *On rings of operators*, Ann. Math., *37* (1936), 116–229.
9. ———, *On rings of operators* IV, Ann. Math., *44* (1943), 716–808.
10. H. Nakano, *Unitärinvarianten hypermaximale normale operatoren*, Ann. Math., *42* (1941), 657–664.
11. I. E. Segal, *Decompositions of operator algebras* I *and* II, Mem. Amer. Math. Soc. No. 9 (1951).

Columbia University and
University of Copenhagen

REDUCIBLE DIOPHANTINE EQUATIONS AND THEIR PARAMETRIC REPRESENTATIONS

E. ROSENTHALL

1. Reducible diophantine equations. The present paper will provide a general method for obtaining the complete parametric representation for the rational integer solutions of the multiplicative diophantine equation

$$1.1 \qquad \prod_{(k)} \prod_{i=1}^{q} \prod_{j=1}^{p} [f_{ki}(x_{1ji}, \ldots, x_{kji})]^{a_{iji}} = \prod_{(k)} \prod_{i=1}^{q} \prod_{j=1}^{p} [h_{ki}(y_{1ji}, \ldots, y_{kji})]^{b_{iji}}$$

for some specified range of k, where the a_{ijk}, b_{ijk} are non-negative integers and the f_{ki}, h_{ki} are decomposable forms, that is to say they are integral irreducible homogeneous polynomials over the rational field R of degree k in k variables which can be written as the product of k linear forms.

Equations of the form 1.1 appear often in diophantine problems and many results concerning their parametric solutions are known. An account of some of these has been given by Skolem **(6**, pp. 64–69**)** where the most general equation of type 1.1 considered is

$$1.2 \qquad f(x_1, x_2, \ldots, x_n) = hy_1^{e_1} y_2^{e_2} \ldots y_p^{e_p},$$

f being a decomposable form of degree n, and we note that 1.1 becomes 1.2 when we restrict the a_{ijk}, b_{ijk} so that $a_{11n} = 1$; $a_{ijk} = 0$ when $i \neq 1$, $j \neq 1$, $k \neq n$; $b_{1j1} = e_j$; $b_{ijk} = 0$ when $i \neq 1$, $k \neq 1$.

The method illustrated there for the complete resolution of 1.2 uses ideal theory in the ring of integers of the algebraic number field K in which f splits, i.e. the field in which f can be written as the product of linear factors. In this process, by solving a simple multiplicative equation **(3**, p. 87**)** in the rational domain, the resolution of 1.2 is reduced to finding all $x_1, x_2, \ldots, x_n$ and $u_1, u_2, \ldots, u_m$ satisfying an equation of the form

$$1.3 \qquad N(\omega_1 x_1 + \omega_2 x_2 + \ldots + \omega_n x_n) = b u_1^{b_1} u_2^{b_2} \ldots u_m^{b_m}$$

with the g.c.d. condition $(x_1, x_2, \ldots, x_n) = 1$ where $\omega_1, \omega_2, \ldots, \omega_n$ is a minimal basis of K and $N(T)$ denotes the norm of the algebraic number T. The resolution of 1.3 is then obtained by factoring each u_i into prime ideals of all permissible degrees and then these prime ideals must be distributed among the linear factors of the left hand member in all possible ways so that these factors are conjugates and the equation is satisfied. No systematic method is provided for making this distribution and although the complete resolution of 1.2 is thus not too difficult in principle the above procedure when applied to 1.2 itself is unwieldy and is certainly unserviceable for the complete resolution of

Received June 14, 1954.

1.1. In view of these circumstances there appears a need for a straightforward method to handle with more facility the general completely reducible equation 1.1.

In this paper we shall consider 1.1 from a different point of view. If f is a decomposable form of degree n then it is well known **(1**, pp. 378–383**)** that a certain integral multiple of f or a form equivalent to f can be expressed as $N(\alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_n x_n)$ where $\alpha_1, \alpha_2, \ldots \alpha_n$ are integers of an algebraic number field of degree n. It follows that equation 1.1 can be replaced by an equation of the form

$$1.4 \qquad a\prod_{(k)}\prod_{i=1}^{q}\prod_{j=1}^{p}[N(\alpha_{1k_i}x_{1ji}+\ldots+\alpha_{kk_i}x_{kji})]^{a_{ij}k} = b\prod_{(k)}\prod_{i=1}^{q}\prod_{j=1}^{p}[N(\beta_{1k_i}y_{1ji}+\ldots+\beta_{kk_i}y_{kji})]^{b_{ij}k}$$

where $\alpha_{rki}, \beta_{rki}$ are algebraic integers of degree k and a, b are rational integers. The resolution of 1.1 thus reduces to solving a multiplicative equation in which each member is a product of linear forms in the normal algebraic extension E/R where E is generated by adjoining to R all the $\alpha_{rki}, \beta_{rki}$ and their conjugates.

The general equation of type 1.4 associated in this form with a given normal extension is formulated below in 2.1 and in the development of our method for the complete resolution of 2.1 it is the Galois group G of the normal extension E which plays a dominant role, and the method permits the equations to be classified according to their group G. We shall show that there is a correspondence between the solutions of 2.1 and the complete parametric representations in multiplicative form of a multiplicative system of independent equations in the rational domain. From a knowledge of these parametric representations, which can be obtained by the method of Bell **(2)** or Ward **(6)**, the complete solution of 2.1 can be written down. The passage from 2.1 to the corresponding multiplicative system in the rational domain and the return from the solutions of this system to the solutions of 2.1 only require the simple and straightforward computation in G as prescribed in Lemma 1. This is established by Theorem 2 and the algorithm in §5.

In an earlier paper the author established Theorem 2 for the case where E is the cyclotomic number field **(4**, p. 219**)**. Although not mentioned in that paper it is readily seen that the same proof shows that this theorem, as stated there, holds for any cyclic field. It was then natural to inquire whether or not a theorem of the same type could be obtained when the Galois group of E is not necessarily cyclic, and to answer this question has led to the present investigation.

2. Notations and conjugate elements. We now introduce some notations and definitions and state a few well-known properties of conjugate elements which are needed in this paper. Let E be a normal algebraic extension of degree n over R and let G be its Galois group; the subgroups of G will be

represented by g, and $o(g)$ stands for the order of g. G can be partitioned into complete conjugate classes $S_1, S_2, \ldots, S_t$ and g_i will denote a representative group from the class S_i. All the small Latin letters will represent rational integers and large Latin letters, unless specified otherwise, will denote ideals in the ring of integers of E. The Greek letters σ, τ, ν are reserved for the substitutions of G. If α is an element of E then $\sigma(\alpha)$, called a conjugate of α, is the image of α under the substitution σ. By σT we mean the ideal obtained from T when we replace each integer α in T by $\sigma(\alpha)$; the ideals σT are called the conjugates of T. All the substitutions of G which leave T unaltered form a group g; we then say that T belongs to g. All the substitutions of G which transform T into a given conjugate element τT form a left coset τg of g. It then follows that all the distinct conjugates of T are given without repetition by $\tau_1 T, \tau_2 T, \ldots, \tau_r T$ where $\tau_1, \tau_2, \ldots, \tau_r$ form a complete set of left residues of G modulo g. Also if T belongs to g then the conjugate ideal σT belongs to $\sigma g \sigma^{-1}$; thus conjugate ideals belong to conjugate subgroups and therefore the only ideals belonging to the set S_j, i.e. to any group of S_j, are the conjugates of ideals belonging to g_j. The letter P will be reserved for prime ideals. $T_{i,k}$ will signify that the ideal T_k belongs to the group g_i.

If $\tau_{i1}, \tau_{i2}, \ldots \tau_{ig(i)}$ be a complete set of left residues of G modulo g_i then with each permissible Galois group G we can associate the multiplicative equation

$$2.1 \qquad \prod_{k=1}^{h} \prod_{i=1}^{t} [\tau_{i1} X_{i,k} \tau_{i2} X_{i,k} \ldots \tau_{ig(i)} X_{i,k}]^{a_{ik}} = \prod_{k=1}^{h} \prod_{i=1}^{t} [\tau_{i1} Y_{i,k} \ldots \tau_{ig(i)} Y_{i,k}]^{b_{ik}}$$

where a_{ik}, b_{ik} are non-negative integers.

Theorem 2 provides a method for the resolution of 2.1 yielding the solution in multiplicative form. If further we select the ideal parameters in this solution of 2.1 so that the ideals X_k and Y_k are principal we obtain the complete solution of 1.4 in rational integers.

3. Lemmas. The following lemmas are required.

LEMMA 1. *Form the following array containing all the left residues of G modulo g_j:*

$$\begin{matrix} \nu_{11,ij}\, \nu_{12,ij} \cdots \nu_{1\,a(1),ij} \\ \nu_{21,ij}\, \nu_{22,ij} \cdots \nu_{1\,a(2),ij} \\ \nu_{s(i,j)1,ij} \cdots \nu_{s(i,j)a(s),ij} \end{matrix}$$

where $\nu_{e1;ij}$ is a left residue modulo g_j not in the first $e-1$ rows and the elements of the eth row are all the distinct left residues modulo g_j of the elements of the right coset $g_i \nu_{e1;ij}$.

Then if A is an ideal belonging to g_j all the distinct conjugates of A are given without repetition by

$$\nu_{ef;ij} A$$

for $e = 1, 2, \ldots, s(i,j); f = 1, 2, \ldots, a(e)$.

This follows since the array contains all the left residues of G modulo g_j and no residue appears twice.

In §6 we shall refer to the above array as the $\nu(i, j)$ array for G.

LEMMA 2. *Let A be unaltered by g_i and let B be the product of all prime ideals belonging to the conjugate set S_j which divide A. Then B can be expressed in the form*

$$\prod_{\tau=1}^{\nu} \prod_{\sigma=1}^{s(i,j)} \left(\prod_{f=1}^{a(\sigma)} \nu_{\sigma f, ij} P_{j,\tau} \right)^{c_{\tau\sigma}},$$

where the ν are as described in Lemma 1 and the $c_{\tau\sigma}$ are non-negative integers.

Proof. Since the only primes belonging to the conjugate set S_j are the conjugates of ideals belonging to g_j then by Lemma 1, B must be of the form

$$B = \prod_{\tau=1}^{\nu} \left(\prod_{\sigma=1}^{s(i,j)} \prod_{f=1}^{a(\sigma)} \nu_{\sigma f, ij} P_{j,\tau} \right)^{d_{\tau\sigma}}.$$

But B is the product of all primes belonging to the same conjugate set which divide an ideal which is unaltered by g_i and so B is also unaltered by g_i. It follows then that if $\nu_{\sigma 1, ij} P_{j,\tau}$ divides B so does $\sigma\nu_{\sigma 1, ij} P_{j,\tau}$ for each $\sigma \in g_i$. However not all these divisors are distinct. The distinct images of $P_{j,\tau}$ by the substitutions of the coset $g_i \nu_{\sigma 1, ij}$ are those obtained by the distinct left residues of $g_i \nu_{\sigma 1, ij}$ modulo g_j, i.e. by the substitutions $\nu_{\sigma 1, ij}, \nu_{\sigma 2, ij}, \ldots, \nu_{\sigma a(\sigma), ij}$. Hence $d_{\tau 1} = d_{\tau 2} = \ldots = d_{\tau a(\sigma)} = c_{\tau\sigma}$, say since by Lemma 1 the conjugates $\nu_{\sigma f, ij} P_{j,\tau}$ are all distinct.

LEMMA 3. *Any left residue of G modulo g_j which appears in the coset τg_i occurs there with multiplicity equal to $o(g_i \cap g_j)$.*

Proof. If σ is in the intersection of the cosets τg_i and σg_j then these cosets both contain $\sigma\nu$ for all $\nu \in g_i \cap g_j$, and these are the only elements in their intersection.

LEMMA 4. *Let $\tau_1, \tau_2, \ldots, \tau_a$ be a complete set of left residues of G modulo g_i and let $\sigma_1, \sigma_2, \ldots, \sigma_b$ be a complete set of left residues of g_i modulo g_j. Then among the products $\tau_r \sigma_s$ $(r = 1, \ldots, a;\ s = 1, \ldots, b)$ appear all the left residues of G modulo g_j and each residue occurs with multiplicity $o(g_j)/o(g_i \cap g_j)$.*

Proof. The cosets $\tau_r g_i$ $(r = 1, 2, \ldots, a)$ contain all the left residues of G modulo g_j each with multiplicity $o(g_j)$. In each coset $\tau_r g_i$ a given left residue modulo g_j which appears there occurs, by Lemma 3, with multiplicity $o(g_i \cap g_j)$. Hence a given left residue of G modulo g_j must appear in $o(g_j)/o(g_i \cap g_j)$ of the cosets $\tau_r g_i$. Since all the residues $\tau\sigma_1, \ldots, \tau\sigma_b$ for given τ are distinct modulo g_j and they are all the distinct left residues of τg_i, the theorem follows.

LEMMA 5. *Let $\tau_1, \ldots, \tau_a$ be a complete set of left residues of G modulo g_i and let $\sigma_1\nu, \ldots, \sigma_b\nu$ for $\sigma \in g_i$ be a complete set of left residues of $g_i\nu$ modulo g_j.*

Then among the products $\tau_r \sigma_s$ $(r = 1, \ldots, a;\ s = 1, \ldots, b)$ *appear all the left residues of* G *modulo* $\nu g_j \nu^{-1}$ *each appearing with multiplicity*

$$o(g_j)/o(g_i \cap \nu g_j \nu^{-1}).$$

This follows from Lemma 4, since $\sigma_1, \sigma_2, \ldots, \sigma_b$ is a complete set of left residues of g_i modulo $\nu g_j \nu^{-1}$.

4. Fundamental theorem. We introduce some additional notation. The notation $A_{j;ik}$ denotes that ideal A_k is the product of all ideals belonging to the conjugate set S_j which divide an ideal unaltered by g_{ik} and $\tau_{i1}, \ldots, \tau_{ig(i)}$ denotes a complete set of left residues of G modulo g_i; $\gamma(m, ij)$ will denote the quotient

$$o(g_j)/o(g_i \cap \nu_{m1,ij}\, g_j\, \nu^{-1}{}_{m1,ij}),$$

where the ν are as prescribed in Lemma 1; also $s(i, j)$, which we shall denote by s in this section, is as defined in Lemma 1.

THEOREM 1. *All solutions of*

$$4.1 \qquad \prod_{k=1}^{h} \prod_{i=1}^{t} (\tau_{i1} A_{j;ik} \cdots \tau_{ig(i)} A_{j;ik})^{a_{ik}} = \prod_{k=1}^{h} \prod_{i=1}^{t} (\tau_{i1} B_{j;ik} \cdots \tau_{ig(i)} B_{j;ik})^{b_{ik}}$$

are given by

$$4.2 \qquad A_{j;ik} = \prod_{q=1}^{w} \prod_{e=1}^{s} (Y_{j,q}^{(e)})^{m_{qe,ik}}, \quad B_{j;ik} = \prod_{q=1}^{w} \prod_{e=1}^{s} (Y_{j,q}^{(e)})^{n_{qe,ik}}$$

where

$$Y_{j,q}^{(e)} = \nu_{e1,ij} T_{j,q} \cdots \nu_{eu(e)} T_{j,q}$$

and

$$m_{q1,ik}, \ldots, m_{qs,ik};\ n_{q1,ik}, \ldots, n_{qs,ik} \qquad (q = 1, 2, \ldots, w)$$

are all the distinct primitive solutions x, y *of the linear system*

$$4.3 \qquad \sum_{k=1}^{h} \sum_{i=1}^{t} \sum_{e=1}^{s} a_{ik} \gamma(e, ij)\, x_{e,ik} = \sum_{k=1}^{h} \sum_{i=1}^{t} \sum_{e=1}^{s} b_{ik} \gamma(e, ij)\, y_{e,ik}.$$

Proof. From Lemma 2 it is seen that $A_{j;ik}$ is of the form

$$4.4 \qquad A_{j;ik} = \prod_{r=1}^{v} \prod_{e=1}^{s} \left(\prod_{f=1}^{a(e)} \nu_{ef,ij} P_{j,r} \right)^{c_{re,ik}}$$

and there is a similar expression for $B_{j;ik}$ with $c_{re,ik}$ replaced by $d_{re,ik}$. Substituting 4.4 in 4.1 gives

$$\prod_{k=1}^{h} \prod_{i=1}^{t} \prod_{r=1}^{v} \prod_{e=1}^{s} \prod_{f=1}^{a(e)} \prod_{b=1}^{g(i)} (\tau_{ib} \nu_{ef,ij} P_{j,r})^{a_{ik} c_{re,ik}} = \prod_{k=1}^{h} \prod_{i=1}^{t} \prod_{r=1}^{v} \prod_{e=1}^{s} \prod_{f=1}^{a(e)} \prod_{b=1}^{g(i)} (\tau_{ib} \nu_{ef,ij} P_{j,r})^{b_{ik} d_{re,ik}}$$

which by Lemma 5 becomes

4.5
$$\prod_{k=1}^{h}\prod_{i=1}^{t}\prod_{r=1}^{v}\prod_{e=1}^{s}[\tau_{j1}P_{j,r}\cdots\tau_{j\theta(j)}P_{j,r}]^{a_{ik}\gamma(e,ij)c_{re,ik}} = \prod_{k=1}^{h}\prod_{i=1}^{t}\prod_{r=1}^{v}\prod_{e=1}^{f}[\tau_{j1}P_{j,r}\cdots\tau_{j\theta(j)}P_{j,r}]^{b_{ik}\gamma(e,ij)d_{re,ik}},$$

and this is equivalent to

$$\prod_{k=1}^{h}\prod_{i=1}^{t}\prod_{r=1}^{v}[\tau_{j1}P_{j,r}\cdots\tau_{j\theta(j)}P_{j,r}]^{a_{ik}c_{r,ik}} = \prod_{k=1}^{h}\prod_{i=1}^{t}\prod_{r=1}^{v}[\tau_{j1}P_{j,r}\cdots\tau_{j\theta(j)}P_{j,r}]^{b_{ik}d_{r,ik}},$$

where

$$c_{r,ik} = \sum_{e=1}^{s} c_{re,ik}\gamma(e,ij), \quad d_{r,ik} = \sum_{e=1}^{s} d_{re,ik}\gamma(e,ij).$$

This relationship is satisfied if and only if

4.6
$$\sum_{k=1}^{h}\sum_{i=1}^{t} a_{ik}c_{r,ik} = \sum_{k=1}^{h}\sum_{i=1}^{t} b_{ik}d_{r,ik} \qquad (r = 1, 2, \ldots, v).$$

This is a linear system of equations to be solved for the non-negative integers $c_{re,ik}, d_{re,ik}$. For fixed r let the w primitive solutions (**5**; p. 9) be

$$m_{q1,ik}, \ldots, m_{qs,ik};\ n_{q1,ik}, \ldots, n_{qs,ik}$$

for $i = 1, 2, \ldots, t;\ k = 1, 2, \ldots, h$ and $q = 1, 2, \ldots, w$. Then all non-negative solutions of 4.6 are given by

4.7
$$c_{re,ik} = \sum_{q=1}^{w} u_{qr}m_{qe,ik}, \quad d_{re,ik} = \sum_{q=1}^{w} u_{qr}n_{qe,ik}$$

for integer parameters u_{qr}. Substituting 4.7 in 4.4 and putting

$$\prod_{r=1}^{v}(P_{j,r})^{u_{qr}} = T_{j,q}$$

we get

$$A_{j;ik} = \prod_{q=1}^{w}\prod_{e=1}^{s}\prod_{f=1}^{a(e)}(\nu_{ef,ij}T_{j,r})^{m_{qe,ik}}, \quad B_{j;ik} = \prod_{q=1}^{w}\prod_{e=1}^{s}\prod_{f=1}^{a(e)}(\nu_{ef,ij}T_{j,r})^{n_{qe,ik}},$$

which we can write as

$$A_{j;ik} = \prod_{q=1}^{w}(Y_{j,q}^{(1)})^{m_{q1,ik}}(Y_{j,q}^{(2)})^{m_{q2,ik}}\cdots(Y_{j,q}^{(s)})^{m_{qs,ik}}$$

and similar expression for $B_{j;ik}$ with $m_{q1,ik}$ replaced by $n_{q1,ik}$.

THEOREM 2. *Let the complete solution in multiplicative form of the equation*

4.8
$$\prod_{k=1}^{h}\prod_{i=1}^{t}\prod_{e=1}^{s}x_{e,ik}^{\gamma(e,ij)a_{ik}} = \prod_{k=1}^{h}\prod_{i=1}^{t}\prod_{e=1}^{s}y_{e,ik}^{\gamma(e,ij)b_{ik}}$$

in the rational domain be

4.9
$$x_{e,ik} = \prod_{q=1}^{w}t_q^{m_{qe,ik}}, \quad y_{e,ik} = \prod_{q=1}^{w}t_q^{n_{qe,ik}} \qquad (e = 1, 2, \ldots, s).$$

Then the complete solution of (4.1) *is given by*

$$A_{j;\alpha} = \prod_{e=1}^{s} x_{e,\alpha}, \quad B_{j;\alpha} = \prod_{e=1}^{s} y_{e,\alpha},$$

where $x_{e,\alpha}, y_{e,\alpha}$ *are as stated in* 4.9 *but with* t_q *replaced by*

$$Y_{j,q}^{(e)} = v_{e1,ij} T_{j,q} \cdots v_{es(e),ij} T_{j,q}.$$

Proof. If 4.9 is the complete solution in multiplicative form of 4.8 then $m_{ee,\alpha}$ and $n_{ee,\alpha}$ are the primitive solutions of equation 4.3 **(6**, p. 70**)**. Thus the values of $A_{j;\alpha}$ and $B_{j;\alpha}$ in 4.1 are precisely the relations 4.2.

5. The algorithm. We can now furnish a procedure for the resolution of the ideal equation

$$5.1 \qquad \prod_{k=1}^{h} \prod_{i=1}^{t} (\tau_{i1} X_{i,k} \tau_{i2} X_{i,k} \cdots \tau_{ip(i)} X_{i,k})^{a_{ik}} = \prod_{k=1}^{h} \prod_{i=1}^{t} (\tau_{i1} Y_{i,k} \tau_{i2} Y_{i,k} \cdots \tau_{ip(i)} Y_{i,k})^{b_{ik}}$$

With $A_{j;\alpha}$ as defined in §4 we can write

$$5.2 \qquad X_{i,k} = \prod_{j=1}^{t} A_{j;\alpha}, \quad Y_{i,k} = \prod_{j=1}^{t} B_{j;\alpha}.$$

Substituting 5.2 into 5.1, and since a given prime ideal belongs to only one conjugate set, we can equate the product of primes belonging to the same conjugate set and we get the system

$$5.3 \qquad \prod_{k=1}^{h} \prod_{i=1}^{t} (\tau_{i1} A_{j;\alpha} \tau_{i2} A_{j;\alpha} \cdots \tau_{ip(i)} A_{j;\alpha})^{a_{ik}} = \prod_{k=1}^{h} \prod_{i=1}^{t} (\tau_{i1} B_{j;\alpha} \tau_{i2} B_{j;\alpha} \cdots \tau_{ip(i)} B_{j;\alpha})^{b_{ik}} \qquad (j = 1, 2, \ldots, t).$$

Each equation of this system is of the type 4.1 and can be solved by the method given there. Substituting the expressions found in this way for $A_{j;\alpha}, B_{j;\alpha}$ into 5.2 gives the complete solution in multiplicative form of the ideal equation 5.1.

Remark. Theorem 2 was stated and proved only for the case when 4.1 and hence the associated system 4.8 consists of two equal products. However the theorem also holds when 4.1 consists of a finite number of equal products. This follows since Ward's result on the correspondence of the solutions of additive and multiplicative equations holds for multiplicative systems with a finite number of equal products.

It will also be seen from the example considered in §6 that the number of parameters in the complete solution of 5.1 obtained here can be reduced. No general results are given here which will yield the solution in terms of a minimum number of parameters.

6. Example. For purposes of illustrating the algorithm we shall consider a very simple equation, one which splits in a field with group G $(1, \sigma)$ where $\sigma^2 = 1$. Representatives from each of the complete conjugate sets are $g_1 = G$, $g_2 = 1$.

Let us consider the problem of solving completely the ideal equation which arises in the complete resolution in rational integers of the equation $x^2 + ay^2 = z^n$, namely,

$$X_{2,1}\cdot\sigma X_{2,1} = Y_{1,1}^n. \tag{6.1}$$

Putting

$$X_{2,1} = A_{1;21}A_{2;21} \text{ and } Y_{1,1} = B_{1;11}B_{2;11}, \tag{6.2}$$

then for equation (6.1) the relations (5.3) become

$$A_{j;21}\cdot\sigma A_{j;21} = B_{j;11}^n \qquad (j = 1, 2),$$

and these equations must be solved for each j. For this purpose we require the $\nu(1, j)$ and $\nu(2, j)$ sets as defined in Lemma 1 and the corresponding multiplicative equation in the rational domain with its parametric solution. This data is listed below for $j = 1, 2$ and the corresponding values of $A_{j;21}, B_{j;11}$ are at once written down.

$j = 1.$ $\nu_{11,11} = 1$; $s(1, 1) = 1, \gamma(1, 11) = 1.$

$\nu_{11,21} = 1$; $s(2, 1) = 1, \gamma(1, 21) = 2.$

$x_{1,21}^2 = y_{1,11}^n \rightarrow x_{1,21} = t_1^n, \quad y_{1,21} = t_1^2.$

$A_{1;21} = T_{1,1}^n \quad B_{1;11} = T_{1,1}^2.$

$j = 2.$ $\nu_{11,12} = 1, \quad \nu_{12,12} = \sigma$; $s(1, 2) = 1, \gamma(1, 12) = 1.$

$\nu_{11,22} = 1$

$\nu_{21,22} = \sigma$; $s(2, 2) = 2, \gamma(1, 22) = 1, \gamma(2, 22) = 1.$

$x_{1,21}x_{2,21} = y_{1,11}^n \rightarrow x_{1,21} = t_1^n t_2^{n-1} \cdots t_n,$

$x_{2,21} = t_2 t_3^2 \cdots t_{n+1}^n,$

$y_{1;11} = t_1 t_2 \cdots t_{n+1}.$

$A_{2;21} = T_{2,1}^n T_{2,2}^{n-1} \cdots T_{2,n}\cdot\sigma(T_{2,2}T_{2,3}^2 \cdots T_{2,n+1}^n),$

$B_{2;11} = T_{2,1}T_{2,2} \cdots T_{2,n+1}\cdot\sigma(T_{2,1}T_{2,2} \cdots T_{2,n+1}).$

Substituting these expression for $A_{j;21}, B_{j;11}$ into (6.2) yields the following for the complete solution of (6.1),

$$X_{2,1} = T_{1,1}^n(T_{2,1}T_{2,2}^{n-1} \cdots T_{2,n})\cdot\sigma(T_{2,2}T_{2,3}^2 \cdots T_{2,n+1}^n),$$

$$Y_{1,1} = T_{1,1}^2(T_{2,1}T_{2,2} \cdots T_{2,n+1})\cdot\sigma(T_{2,1}T_{2,2} \cdots T_{2,n+1}).$$

In conclusion we list the result obtained when the method of this paper is applied to an equation which splits in a field whose group G is the group of

symmetries of the square. G is then a group of order eight generated by σ, τ where $\tau^2 = \sigma^4 = 1$, $\tau\sigma = \sigma^3\tau$, $\tau\sigma^2 = \sigma^2\tau$, $\tau\sigma^3 = \sigma\tau$. Representatives from each of the complete conjugate sets can be taken to be $g_1 = G, g_2 = (1, \sigma, \sigma^2, \sigma^3)$, $g_3 = (1, \sigma^2, \tau, \sigma^2\tau)$, $g_4 = (1, \sigma^2, \sigma\tau, \sigma^3\tau)$, $g_5 = (1, \sigma^2)$, $g_6 = (1, \tau)$, $g_7 = (1, \sigma\tau)$, $g_8 = \tau$.

Then by the method of this paper the complete solution of the ideal equation

$$X_{6,1} \cdot \sigma X_{6,1} \cdot \sigma^2 X_{6,1} \cdot \sigma^3 X_{6,1} = U_{2,1} \cdot \tau U_{2,1} \tag{6.3}$$

is found to be

$$X_{6,1} = N_{2,3}^{1+\tau} N_{3,2} N_{8,1}^{1+\tau}$$

$$U_{2,1} = N_{2,3}^{3+\tau} N_{3,2}^{1+\sigma} N_{8,1}^{1+\sigma+\sigma^2+\sigma^3},$$

where the notation $T^{a\sigma+b\tau}$ is used for $\sigma T^a \cdot \tau T^b$. Equation 6.3 arises, for example, in the complete resolution in rational integers of the equation

$$N(a + b\theta + c\theta^2 + d\theta^3) = u^2 + v^2$$

where $\theta = \sqrt[4]{a}$.

References

1. P. Bachman, *Die Arithmetik der quadratischen Formen* (Berlin, 1923).
2. E. T. Bell, *Reciprocal arrays and diophantine analysis*, Amer. J. Math., *55* (1933), 50–66.
3. ———, *Separable diophantine equations*, Trans. Amer. Math. Soc., *57* (1945), 86–101.
4. E. Rosenthall, *Diophantine equations separable in cyclotomic fields*, Duke Math. J. *20* (1953), 141–338.
5. T. Skolem, *Diophantische Gleichungen*, Ergebnisse der Math. und ihrer Grenzgebiete, *5*, no. 4 (Berlin, 1938).
6. Morgan Ward, *A type of multiplicative diophantine system*, Amer. J. Math., *55* (1933), 67–76.

McGill University

AN INHOMOGENEOUS MINIMUM FOR NON-CONVEX STAR-REGIONS WITH HEXAGONAL SYMMETRY

R. P. BAMBAH AND K. ROGERS

1. Introduction. Several authors have proved theorems of the following type:

Let x_0, y_0 be any real numbers. Then for certain functions $f(x, y)$, there exist numbers x, y such that

1.1 $$x \equiv x_0, \quad y \equiv y_0 \pmod 1,$$

and

1.2 $$|f(x, y)| < \max\,[|f(\tfrac{1}{2}, 0)|, |f(0, \tfrac{1}{2})|, |f(\tfrac{1}{2}, \tfrac{1}{2})|, |f(\tfrac{1}{2}, -\tfrac{1}{2})|].$$

The first result of this type, but with $|f(\frac{1}{2}, \frac{1}{2})|, |f(\frac{1}{2}, -\frac{1}{2})|$ replaced by $\min |f(\frac{1}{2}, \pm\frac{1}{2})|$, was given by Barnes **(3)** for the case when the function is an indefinite binary quadratic form. A generalisation of this was proved by elementary geometry by K. Rogers **(6)**. Bambah **(1)** proved the theorem for binary cubic forms with three real linear factors, and Chalk **(4)** proved the same result for binary cubic forms with only one real linear factor. Mordell **(5)** generalised Chalk's result and proved that for functions $f(x, y)$ satisfying certain conditions, including the condition

1.3 $$|f(x, y)| < k|f(2x, 2y)|,$$

for some k independent of x, y, one can find x, y to satisfy 1.1 and also

1.4 $$|f(x, y)| < k.\ \max\,[|f(1, 0)|, |f(0, 1)|, |f(1, 1)|, |f(1, -1)|].$$

Any function satisfying 1.3 and 1.2 also satisfies 1.4, and in fact one can modify Mordell's proof very slightly to get the theorem in the form 1.2 without imposing a condition 1.3. It is only when the function is not homogeneous that the results differ.

Since Mordell's and Rogers' papers were elementary generalisations to certain regions with one and two asymptotes respectively of the results of Chalk and Barnes, it might be interesting to see what properties of a region $f(x, y) < 1$ with three asymptotes through the origin are necessary in order that the result 1.2 may be proved. In this way, the essential property of the binary cubic required in Bambah's theorem is revealed, namely that the region can be transformed by a linear transformation into one with hexagonal symmetry.

2. Equivalent forms of the theorem, and some lemmas. Let l_1Ol_4, l_2Ol_5, l_3Ol_6 be three lines through the origin O such that Ol_2, Ol_3 make angles of 60°

Received September 27, 1954. The authors wish to thank Professor Mordell, whose criticism and advice have improved this paper considerably.

and 120° respectively with Ol_1. Bambah (2) defines a star-region $\mathfrak{R}$ as having *hexagonal symmetry* if

(i) $\mathfrak{R}$ is symmetric with respect to the lines $l_1Ol_4, l_2Ol_5, l_3Ol_6$ and their bisectors,

(ii) the boundary $\mathfrak{B}$ of $\mathfrak{R}$ either terminates in the lines $l_1Ol_4, \ldots, l_3Ol_6$ or has them as asymptotes,

(iii) the region external to $\mathfrak{R}$ and lying between Ol_1 and Ol_2 is convex,

(iv) each of the six branches of $\mathfrak{B}$ is a continuous curve.

The arithmetical form of the result to be proved is the following.

THEOREM 1. *Suppose the region $f(x, y) \leqslant 1$ has hexagonal symmetry, and let $\alpha, \beta, \gamma, \delta$ be fixed real numbers with $\alpha\delta - \beta\gamma \neq 0$. Then for any real u_0, v_0 there exist numbers $(u, v) \equiv (u_0, v_0) \pmod 1$, such that*

$$f(\alpha u + \beta v, \gamma u + \delta v) \leqslant \max\left[f\left(\frac{\alpha}{2}, \frac{\gamma}{2}\right), f\left(\frac{\beta}{2}, \frac{\delta}{2}\right), f\left(\frac{\alpha \pm \beta}{2}, \frac{\gamma \pm \delta}{2}\right)\right].$$

It is usually convenient to state the theorem geometrically. Let Λ be the lattice generated by the points $A(\alpha, \gamma)$ and $B(\beta, \delta)$, and let $C = A + B$. Let $\mathfrak{R}$ denote the region given by

$$f(x, y) \leqslant \max\left[f\left(\frac{\alpha}{2}, \frac{\gamma}{2}\right), f\left(\frac{\beta}{2}, \frac{\delta}{2}\right), f\left(\frac{\alpha \pm \beta}{2}, \frac{\gamma \pm \delta}{2}\right)\right].$$

Then $\mathfrak{R}$ is a region with hexagonal symmetry which contains the points $\pm\frac{1}{2}A, \pm\frac{1}{2}B, \pm\frac{1}{2}(A \pm B)$. We denote by $\mathfrak{R}+P$ the region obtained from $\mathfrak{R}$ by the translation which moves O to P. Then the above theorem is equivalent to the following.

THEOREM 2. *The parellelogram $OACB$, and hence the whole plane, is covered completely by the regions $\mathfrak{R} + P, P \in \Lambda$.*

Let Ω be any automorph of $\mathfrak{R}$, and write $\Omega\Lambda = \Lambda', \Omega A = A', \Omega B = B'$. Then $\mathfrak{R}$ contains $\pm\frac{1}{2}A', \pm\frac{1}{2}B', \pm\frac{1}{2}(A' \pm B')$, and the plane is covered by $\mathfrak{R} + \Lambda$ if and only if it is covered by $\mathfrak{R} + \Lambda'$. There is no loss of generality if we choose co-ordinates so that one asymptote is the x-axis. Then, since rotations through 60° or reflections in the axes are automorphisms of $\mathfrak{R}$, we can suppose by a suitable choice of Ω that one of the pairs $\Omega A, \Omega B$; $\Omega B, \Omega A$; $\Omega A, \Omega(-B)$; $\Omega(-B), \Omega A$, which we can still call (A, B), satisfies

(1) $OA \leqslant OB$,

(2) the angle AOB is acute,

(3) the rotation from OA to OB is anti-clockwise,

(4) A lies in the region $0 \leqslant y \leqslant x\sqrt{3}$.

Then we have only to prove the following:

Let Λ be a lattice generated by points A, B satisfying (1), (2), (3), (4). *Let $\mathfrak{R}$ be a region, say $f(x, y) \leqslant k$, with hexagonal symmetry which contains the points $\frac{1}{2}A, \frac{1}{2}B, \frac{1}{2}(A \pm B)$, and for which the x-axis is an asymptote. Then the regions $\mathfrak{R} + \Lambda$ cover the parallelogram $OACB$ and hence the plane.*

For the proof we need three lemmas:

(a) *Let $PQRS$ be a parallelogram with PQ and PS parallel to Ol_i and Ol_{i+1} respectively ($i = 1, \ldots, 6$; $Ol_7 = Ol_1$). Define $\mathfrak{R}'$ by $f(x, y) \leqslant f(\alpha, \beta)$, where α, β are the co-ordinates of the point $\frac{1}{2}(R - P)$. Then $PQRS$ is covered by $\mathfrak{R}' + P, \mathfrak{R}' + R$.*

The region $\mathfrak{R}' + P$ has as part of its boundary an arc which passes through the point $\frac{1}{2}(R - P) + P = \frac{1}{2}(R + P)$ and has PQ, PS as asymptotes. For the region $\mathfrak{R}' + R$, the corresponding point and asymptotes are $-\frac{1}{2}(R - P) + R = \frac{1}{2}(P + R)$ and RS, RQ. The two arcs have a common tac-line at the point $\frac{1}{2}(P + R)$ and so do not cross in the parallelogram $PQRS$. Hence every point of the parallelogram is covered by one or other of the two regions, the small regions near Q, S in fact being covered twice.

(b) *Let PQR be an equilateral triangle with sides parallel to Ol_1, Ol_2, Ol_3 in some order. Let $\mathfrak{R}'$ be the region $f(x, y) \leqslant f(\frac{1}{2}\gamma\sqrt{3}, \frac{1}{2}\gamma)$, where γ is the distance of P from QR. Then PQR is covered by $\mathfrak{R}' + P$.*

Since the point $(\frac{1}{2}\gamma\sqrt{3}, \frac{1}{2}\gamma)$ lies on the bisector of the angle l_1Ol_2, a tac-line there to the boundary of $\mathfrak{R}'$ is parallel to Ol_3. Hence the boundary of $\mathfrak{R}' + P$ has an arc which passes through the foot of the perpendicular from P to QR, touches QR at this point, and has PQ, PR as asymptotes; and so PQR is covered.

(c) *If $0 \leqslant y_1 \leqslant x_1\sqrt{3}$, and $x_1 \leqslant x_2$, then $f(x_1, y_1) \leqslant f(x_2, y_1)$.*

This follows immediately from the convexity and the relationship of the region $\mathfrak{R}$ to the x-axis.

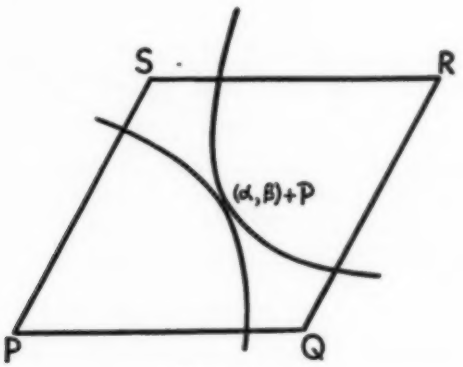

Figure 1

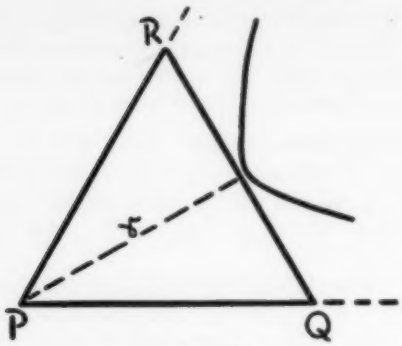

Figure 2

3. Proof of the theorem. The conditions (1) to (4) on A, B imply that A lies between Ol_1 and Ol_2, and B lies between Ol_i and Ol_{i+1}, where $i = 1, 2$ or 3. When B lies between Ol_2 and Ol_3, denote the point of intersection of Ol_2 with the line through B parallel to Ol_4 by F and that of Ol_2 with the line through A parallel to Ol_4 by G. Taking Ol_1 to be the positive x-axis, we have the following cases to consider.

(i) B lies between Ol_1 and Ol_2.

(ii) B lies between Ol_2 and Ol_3, and F lies above G, in the sense that the ordinate of F is not less than the ordinate of G.

(iii) B lies between Ol_2 and Ol_3, and F lies below G.

(iv) B lies between Ol_3 and Ol_4, and B is above A.

(v) B lies between Ol_3 and Ol_4, and B is below A.

Let the co-ordinates of A and B be (p, q) and (r, s) respectively.

(i)

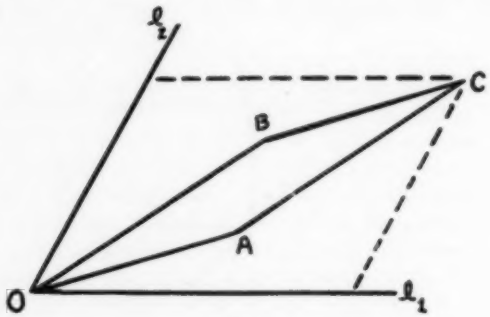

Figure 3

By (a), the parallelogram bounded by the lines through O and C parallel to Ol_1, Ol_2 is covered by $\mathfrak{R}$, $\mathfrak{R} + C$, and hence $OACB$ is covered.

(ii)

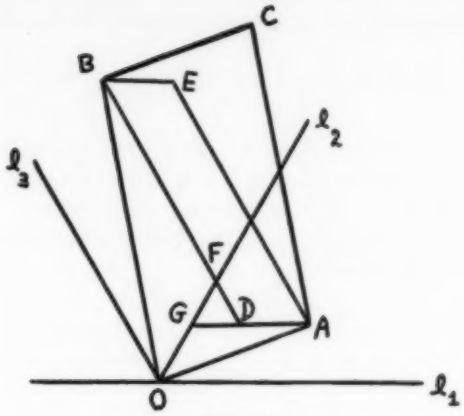

Figure 4

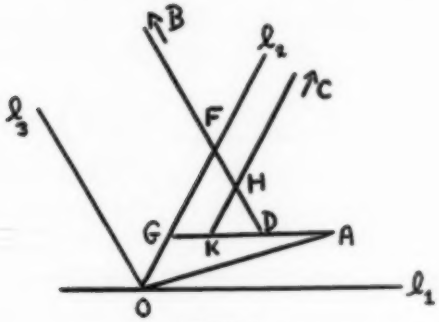

Figure 5

Since $OA \leqslant OB$, and since B lies in the second sector while A lies in the first, we have

$$y_A \leqslant OA \sin \tfrac{1}{3}\pi \leqslant OB \sin \tfrac{1}{3}\pi \leqslant y_B,$$

and so B lies above A. Draw lines ADG, AE parallel to Ol_4 and Ol_3, and lines BFD and BE parallel to Ol_3 and Ol_1 respectively. Since F lies above G, it is clear that D is in the first sector. The treatment differs according as C is in the second or first sector, the corresponding figures being respectively (4) and (5), but we do not separate into cases until it is necessary.

By (a), we see that $ADBE$ is covered by $\mathfrak{N} + A$, $\mathfrak{N} + B$, OAG by $\mathfrak{N}$, $\mathfrak{N} + A$, and OFB by $\mathfrak{N}$, $\mathfrak{N} + B$. Because of the symmetry of $OACB$ about its centre, it will now be enough to prove that the triangle DFG is covered by $\mathfrak{N}$, $\mathfrak{N} + A$, $\mathfrak{N} + B$, $\mathfrak{N} + C$.

Since the distances of DF from O and FG from A are $\frac{1}{2}(s + r\sqrt{3})$ and $\frac{1}{2}(p\sqrt{3} - q)$ respectively, it follows from (b) that DFG is covered by each of the sets

$$f(x, y) \leqslant f\{\tfrac{1}{4}\sqrt{3}(s + r\sqrt{3}), \tfrac{1}{4}(s + r\sqrt{3})\},$$

and

$$[f(x, y) \leqslant f\{\tfrac{1}{4}\sqrt{3}(p\sqrt{3} - q), \tfrac{1}{4}(p\sqrt{3} - q)\}] + A.$$

Now since B lies between Ol_2 and Ol_3, we have $0 \leqslant |r|\sqrt{3} \leqslant s$, and we take in turn the cases when r is negative and when r is positive:

(I) $0 \leqslant -r\sqrt{3} \leqslant s$: by reflection in Oy, then rotation through $\frac{1}{3}\pi$ in the clockwise direction, we deduce in turn that

$$\begin{aligned} f(\tfrac{1}{2}r, \tfrac{1}{2}s) &= f(-\tfrac{1}{2}r, \tfrac{1}{2}s) \\ &= f\left(\frac{-r + s\sqrt{3}}{4}, \frac{s + r\sqrt{3}}{4}\right) \\ &\geqslant f(\tfrac{1}{4}\sqrt{3}(s + r\sqrt{3}), \tfrac{1}{4}(s + r\sqrt{3})), \end{aligned}$$

by (c), since for $r \leqslant 0$ we have $s\sqrt{3} - r \geqslant s\sqrt{3} + 3r$. Hence, in this case DFG is covered by the region $f(x, y) \leqslant f(\frac{1}{2}r, \frac{1}{2}s)$ and so certainly by $\mathfrak{R}$.

(II) $0 \leqslant r\sqrt{3} \leqslant s$: by rotation through $\frac{1}{3}\pi$ in the clockwise direction, we see that

$$\begin{aligned} f(\tfrac{1}{2}r, \tfrac{1}{2}s) &= f\left(\frac{r + s\sqrt{3}}{4}, \frac{s - r\sqrt{3}}{4}\right) \\ &\geqslant f(\tfrac{1}{4}\sqrt{3}(s - r\sqrt{3}), \tfrac{1}{4}(s - r\sqrt{3})), \end{aligned}$$

the inequality following from (c), since for $r \geqslant 0$ we have $r + s\sqrt{3} \geqslant s\sqrt{3} - 3r$. Thus, *if we have* $s - r\sqrt{3} \geqslant p\sqrt{3} - q$, we can conclude that

$$f(\tfrac{1}{2}r, \tfrac{1}{2}s) \geqslant f(\tfrac{1}{4}\sqrt{3}(p\sqrt{3} - q), \tfrac{1}{4}(p\sqrt{3} - q)),$$

and so, by the remarks preceding (I), we see that DFG is covered by $\mathfrak{R} + A$.

Now suppose that $s - r\sqrt{3} < p\sqrt{3} - q$. This means that C is in the first sector, since the distance of A from the line through C parallel to Ol_3 is $\frac{1}{2}(s - r\sqrt{3})$, while the distance of A from Ol_2 is $\frac{1}{2}(p\sqrt{3} - q)$. The figure is as in Figure 5 and there are two possibilities:

Either, the triangle DFG lies completely between Ol_2 and the line through C parallel to Ol_3 and is consequently covered by $\mathfrak{R}$, $\mathfrak{R} + C$;

Or, the line through C parallel to Ol_3 cuts the lines DF, DG at points H, K, thus dividing DFG into the triangle DHK and the trapezium $FGKH$. Of these, $FGKH$ is covered by $\mathfrak{R}$, $\mathfrak{R} + C$, and DKH is covered by

$$[f(x, y) \leqslant f\{\tfrac{1}{4}\sqrt{3}(s - r\sqrt{3}), \tfrac{1}{4}(s - r\sqrt{3})\}] + A,$$

since $\frac{1}{2}(s - r\sqrt{3})$ is the distance of HK from A. Hence, by the inequality above, we see that DKH is covered by $\mathfrak{R} + A$. This completes the investigation of case (ii).

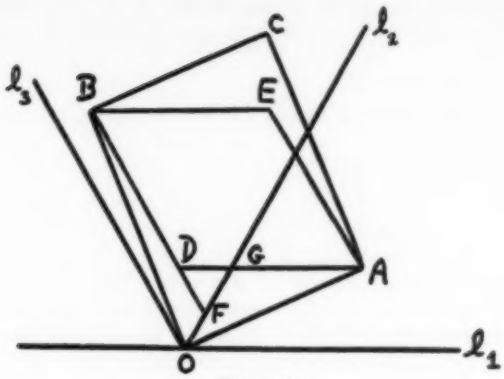

Figure 6

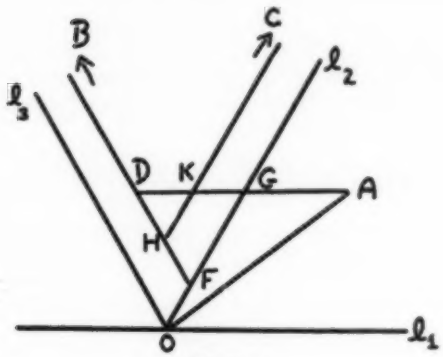

Figure 7

(iii) In this case, as F lies below G, D is in the second sector. Figures 6 and 7 represent the situation when the line through C parallel to Ol_3 does not meet or does meet the triangle DFG. The case when C is in the first sector is not drawn but will be considered. As in (ii), we have only to show that DFG is covered by $\mathfrak{N}$, $\mathfrak{N}+A$, $\mathfrak{N}+B$, $\mathfrak{N}+C$.

Since the perpendicular distances from O to DG and from B to GF are q and $\frac{1}{2}(s-r\sqrt{3})$ respectively, it is clear that DFG is covered by each of the sets

$$f(x, y) \leqslant f(\tfrac{1}{2}\sqrt{3}q, \tfrac{1}{2}q),$$

and

$$[f(x, y) \leqslant f\{\tfrac{1}{4}(s-r\sqrt{3})\sqrt{3}, \tfrac{1}{4}(s-r\sqrt{3})\}] + B$$

Now since A lies in the first sector, we have $O \leqslant q \leqslant p\sqrt{3}$, but in fact we divide into cases according as $q\sqrt{3}$ is not greater than or not less than p.

(I) $q\sqrt{3} < p$. In this case, (c) implies that

$$f(\tfrac{1}{2}p, \tfrac{1}{2}q) > f(\tfrac{1}{2}\sqrt{3}q, \tfrac{1}{2}q),$$

and hence DFG is covered by $\mathfrak{R}$.

(II) $q\sqrt{3} > p$. By reflection in the line $x = y\sqrt{3}$, the bisector of l_2Ol_1, we have

$$\begin{aligned} f(\tfrac{1}{2}p, \tfrac{1}{2}q) &= f(\tfrac{1}{4}(p + q\sqrt{3}), \tfrac{1}{4}(p\sqrt{3} - q)) \\ &> f\left(\frac{\sqrt{3}(p\sqrt{3} - q)}{4}, \frac{(p\sqrt{3} - q)}{4}\right), \end{aligned}$$

using (c) at the second stage, since $q\sqrt{3} > p$ implies that $p+q\sqrt{3} > 3p-q\sqrt{3}$. Hence, using the second region above which we showed covered DFG, we see that DFG is covered by $\mathfrak{R} + B$ if we make the further assumption that $p\sqrt{3} - q > s - r\sqrt{3}$. This leaves the cases when $p\sqrt{3} - q < s - r\sqrt{3}$. In this case B is nearer to the line through C parallel to Ol_3 than to OFG, so that certainly C is in the second sector. Figure 6 indicates the case when DFG lies entirely between Ol_2 and the line through C parallel to Ol_3; in this case, as before, DFG is covered by $\mathfrak{R}$, $\mathfrak{R} + C$. The case when the line through C parallel to Ol_3 meets DFG is shown in Figure 7, where it is seen that the triangle is divided into the trapezium $FGKH$ and the triangle DHK. The trapezium is covered by $\mathfrak{R}$, $\mathfrak{R} + C$, while DHK is covered by

$$\left[f(x, y) < f\left(\frac{\sqrt{3}(p\sqrt{3} - q)}{4}, \frac{p\sqrt{3} - q}{4}\right)\right] + B,$$

since the distance of B from HK is $\frac{1}{2}(p\sqrt{3} - q)$. By the inequality proved under (II), we infer that DHK is covered by

$$[f(x, y) < f(\tfrac{1}{2}p, \tfrac{1}{2}q)] + B,$$

and hence, a fortiori, is covered by $\mathfrak{R} + B$.

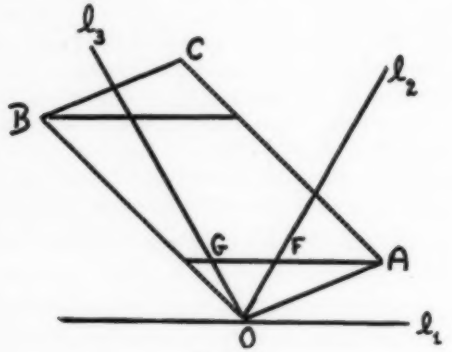

Figure 8

(iv) By applications of (a), we have only to show that the triangle OFG is covered, as shown in Figure 8. Since the distance from B to OF is $\frac{1}{2}(s - r\sqrt{3})$, OFG is covered by

$$\left[f(x, y) \leqslant f\left(\frac{\sqrt{3}(s - r\sqrt{3})}{4}, \frac{s - r\sqrt{3}}{4}\right)\right] + B.$$

Hence, if $s - r\sqrt{3} \leqslant p\sqrt{3} - q$, then OFG is covered by

$$\left[f(x, y) \leqslant f\left(\frac{\sqrt{3}(p\sqrt{3} - q)}{4}, \frac{p\sqrt{3} - q}{4}\right)\right] + B;$$

and then, since the angle AOB is acute, A lies above the line $y = x/\sqrt{3}$, hence $q\sqrt{3} \geqslant p$, hence, as in (II) of (iii),

$$f\left(\frac{\sqrt{3}(p\sqrt{3} - q)}{4}, \frac{p\sqrt{3} - q}{4}\right) \leqslant f(\tfrac{1}{2}p, \tfrac{1}{2}q),$$

and therefore OFG is covered by $\mathfrak{N} + B$. Finally, suppose that $p\sqrt{3} - q < s - r\sqrt{3}$. Then C is in the second sector and, as before, we can show that OFG is covered by $\mathfrak{N}$, $\mathfrak{N} + C$, or by $\mathfrak{N}$, $\mathfrak{N} + C$, and

$$\left[f(x, y) \leqslant f\left(\frac{\sqrt{3}(p\sqrt{3} - q)}{4}, \frac{p\sqrt{3} - q}{4}\right)\right] + B.$$

Using the same inequality as above, we deduce the result.

Case (v) is similar to (iv). In fact, since the relation $OA \leqslant OB$ does not play any part in (iv), one can use the same proof after reflection in the y-axis.

This completes the proof of Theorem 1. We note that the argument also shows that, if the regions are strictly convex, then strict inequality can be obtained in the statement of the theorem, except possibly when (u_0, v_0) is congruent to one of $(\frac{1}{2}, 0)$, $(0, \frac{1}{2})$, $(\frac{1}{2}, \frac{1}{2})$.

4. Concluding remarks. The example

$$f(x, y) = |xy(x - y)(x + y)|$$

shows that the results for regions with one, two or three asymptotes do not have an analogue for general regions with four or more asymptotes. This is clear from the fact that for the above function the right-hand side of (1.2) is zero.

References

1. R. P. Bambah, *Non-homogeneous binary cubic forms*, Proc. Camb. Phil. Soc., *47* (1951), 457–460.
2. R. P. Bambah, *On the geometry of numbers of non-convex star-regions with hexagonal symmetry*, *Phil. Trans. Royal Soc.* A *243* (1951), 431–462.
3. E. S. Barnes, *Non-homogeneous binary quadratic forms*, Quarterly J. Math. (2), *1* (1950), 199–210.
4. J. H. H. Chalk, *The minimum of a non-homogeneous binary cubic form*, Proc. Camb. Phil. Soc. *48* (1952), 392–401.
5. L. J. Mordell, *The minima of some inhomogeneous functions of two variables*, Duke Math. J. *19* (1952), 519–527.
6. K. Rogers, *The minima of some inhomogeneous functions of two variables*, J. Lond. Math. Soc. *28* (1953), 394–402.

Punjab University *Princeton University*

THE DISTRIBUTION OF TOTATIVES

D. H. LEHMER

1. Introduction. This paper is concerned with the numbers which are relatively prime to a given positive integer

$$n = p_1^{a_1} p_2^{a_2} \dots p_t^{a_t}$$

where the p's are the distinct prime factors of n. Since these numbers recur periodically with period n, it suffices to study the $\phi(n)$ numbers $\leqslant n$ and relatively prime to n. Here

$$\phi(n) = n(1 - p_1^{-1})(1 - p_2^{-1}) \dots (1 - p_t^{-1}) \tag{1}$$

is Euler's function. Following Sylvester, these $\phi(n)$ numbers are called the *totatives* of n. One may ask how these totatives are distributed among the integers $\leqslant n$. Specifically we may divide the interval from 0 to n into k equal subintervals and consider the number of totatives in each of these subintervals. It is natural to suppose that these intervals are of more than unit length so that we shall suppose that $n > k$ in what follows. The ambiguity of assigning an interval to a totative which occupies the common end point of two adjacent intervals does not arise. In fact if qn/k is a totative, n must divide k. But $n > k$. Hence we define, for each $q = 0, 1, \dots, k - 1$, the partial totient function $\phi(k, q, n)$ as the number of totatives r for which

$$nq/k < r < n(q + 1)/k. \tag{2}$$

Alternatively, one may divide the unit circle into k equal arcs by the kth roots of unity and enquire about the number of *primitive* nth roots of unity in each such arc. It is clear that

$$\sum_{q=0}^{k-1} \phi(k, q, n) = \phi(1, 0, n) = \phi(n). \tag{3}$$

We shall be interested in the question of how uniformly the totatives are distributed and so we introduce the function

$$E(k, q, n) = \phi(n) - k\phi(k, q, n), \tag{4}$$

which may be described as the excess of the number of all totatives of n over the number there would be if the totatives were everywhere as dense as they are in the interval

$$nq/k \leqslant x \leqslant n(q + 1)/k. \tag{5}$$

The value of $E(k, q, n)$ is an integer, positive, negative, or zero. By (3) we have

$$\sum_{q=0}^{k-1} E(k, q, n) = 0. \tag{6}$$

Received November 5, 1954.

2. Uniform distribution. The vanishing of $E(k, q, n)$ is an indication of uniformly distributed totatives. For $E(k, q, n)$ to vanish even for one value of q it is necessary that $\phi(n)$ be divisible by k. That this condition is not sufficient is seen from the example of $n = 21$ and $k = 4$.

In this case $\phi(n) = 12$ is divisible by k but, as we shall see,

$$E(4, 0, 21) = E(4, 3, 21) = -4,$$
$$E(4, 1, 21) = E(4, 2, 21) = 4,$$

so that $E(4, q, 21)$ never vanishes.

If, for some k and n, the functions $E(k, q, n)$ vanish for all values of q, then we say that the totatives of n are uniformly distributed with respect to k, there being $\phi(n)/k$ totatives in each of the k intervals. For example, for every $n > 2$ the totatives are uniformly distributed with respect to $k = 2$. That is

$$E(2, 0, n) = E(2, 1, n) = 0 \qquad (n > 2).$$

This follows at once from the fact that if τ is a totative, so also is $n - \tau$. Similarly we have

THEOREM 1. *If $n > k$, $E(k, q, n) = E(k, k - q - 1, n)$ for all values of q.*

THEOREM 2. *If n is divisible by k^2 then the totatives of n are uniformly distributed with respect to k, that is,*

$$E(k, q, hk^2) = 0 \qquad (q = 0, 1, \ldots, k-1).$$

Proof. This follows at once if we consider the fact that all the totatives of $n = hk^2$ may be generated from those less than hk by adding successive multiples of hk. In fact the integers

$$\tau + qhk \qquad (q = 0, 1, \ldots, k - 1; 0 < \tau < hk)$$

are all totatives of $n = hk^2$ if τ is, and every totative of n is of this form.

3. Auxiliary numerical functions. We proceed to develop formulas for $E(k, q, n)$ in terms of simpler numerical functions. These functions are:

$\mu(n)$, Möbius' function;
$\lambda(n)$, Liouville's function;
$\theta(n)$, the number of square-free divisors of n.

All these functions, as well as $\phi(n)$, are multiplicative, that is, if f is any one of these functions, then

$$f(n) = \prod_{i=1}^{t} f(p_i^{\alpha_i}).$$

For prime power arguments we have:

$$\mu(p^\alpha) = \begin{cases} -1 & \text{if } \alpha = 1, \\ 0 & \text{if } \alpha > 1 \end{cases}$$
$$\lambda(p^\alpha) = (-1)^\alpha,$$
$$\theta(p^\alpha) = 2.$$

In particular,

$$\theta(n) = 2^t = \sum_{\delta \mid n,\ \mu(\delta) \neq 0} 1.$$

We conclude this list of well-known facts by quoting the following formulas:

$$\sum_{\delta \mid n} \mu(n/\delta) = 0, \quad n > 1, \tag{7}$$

$$\sum_{\delta \mid n} \lambda(\delta)\, \mu(n/\delta) = \lambda(n)\, \theta(n), \tag{8}$$

where the sums, as indicated, range over all the divisors δ of n. The second of these is due to Liouville **(3)**.

In what follows we denote by $[x]$ the greatest integer $\leqslant x$.

THEOREM 3.

$$E(k, q, n) = \sum_{\delta \mid n} \left\{ \delta + k\left[\frac{q\delta}{k}\right] - k\left[\frac{(q+1)\delta}{k}\right] \right\} \mu(n/\delta).$$

Proof. By a theorem of Legendre **(1**, pp. 7–8**)** the number of totatives of n which do not exceed x is given by

$$\sum_{\delta \mid n} [x/\delta]\, \mu(\delta) = \sum_{\delta \mid n} [\delta x/n]\, \mu(n/\delta).$$

In particular,

$$\phi(n) = \sum_{\delta \mid n} \delta\, \mu(n/\delta) \tag{9}$$

and

$$\phi(k, q, n) = \sum_{\delta \mid n} \{[\delta(q+1)/k] - [\delta q/k]\}\, \mu(n/\delta). \tag{10}$$

The theorem now follows from (4).

For $q = 0$ we have the simple formula

$$E(k, 0, n) = \sum_{\delta \mid n} r_k(\delta)\, \mu(n/\delta), \tag{11}$$

where $r_k(\delta)$ is the least positive remainder, $\delta - k[\delta/k]$, on division of δ by k.

THEOREM 4. *If n is divisible by a prime p of the form $kx + 1$, then the totatives of n are uniformly distributed with respect to k, that is,*

$$E(k, q, n) = 0 \qquad (q = 0, 1, \ldots, k-1).$$

Proof. It suffices to show that in case

$$n = p^a m, \quad p = kx + 1, \quad p \nmid m,$$

then $\phi(k, q, n)$ is not a function of q. Now

$$\phi(k, q, n) = \sum_{\delta \mid m} \mu(m/\delta)\, g(k, q, p^a, \delta), \tag{12}$$

where

$$g(k, q, p^a, \delta) = \sum_{v=0}^{a} \left\{ \left[\frac{p^v \delta(q+1)}{k}\right] - \left[\frac{p^v \delta q}{k}\right] \right\} \mu(p^{a-v}).$$

Since $\mu(p^\beta) = 0$ for $\beta > 1$, we have

$$g(k, q, p^\alpha, \delta) = \left[\frac{p^\alpha\delta(q+1)}{k}\right] - \left[\frac{p^\alpha\delta q}{k}\right] - \left[\frac{p^{\alpha-1}\delta(q+1)}{k}\right] + \left[\frac{p^{\alpha-1}\delta q}{k}\right].$$

Let $p^\alpha = kr + 1$, $p^{\alpha-1} = ks + 1$. Then:

$$\left[\frac{p^\alpha\delta q}{k}\right] = r\delta q + \left[\frac{\delta q}{k}\right],$$

$$\left[\frac{p^{\alpha-1}\delta q}{k}\right] = s\delta q + \left[\frac{\delta q}{k}\right],$$

$$\left[\frac{p^\alpha\delta(q+1)}{k}\right] = r\delta q + r\delta + \left[\frac{\delta(q+1)}{k}\right],$$

$$\left[\frac{p^{\alpha-1}\delta(q+1)}{k}\right] = s\delta q + s\delta + \left[\frac{\delta(q+1)}{k}\right].$$

Substituting, we find

$$g(k, q, p^\alpha, \delta) = (r - s)\delta = \delta\,\phi(p^\alpha)/k.$$

Since this is not a function of q, the theorem follows.

Incidentally we may substitute the value obtained for $g(k, q, p^\alpha, \delta)$ into (12) and get

$$\phi(k, q, n) = \sum_{\delta|m} \mu(m/\delta)\,\delta\,\frac{\phi(p^\alpha)}{k} = \phi(p^\alpha)\,\frac{\phi(m)}{k} = \frac{\phi(n)}{k},$$

as it should be.

We define n as an "exceptional number with respect to k" in case n is divisible either by k^2 or by a prime of the form $kx + 1$. Theorems 2 and 4 together state that if n is an exceptional number with respect to k, then the totatives are uniformly distributed with respect to k. We may confine our attention in what follows to non-exceptional numbers. Every number is exceptional with respect to 2. Hence we consider $k \geqslant 3$.

The cases $k = 3, 4, 6$, are sufficiently simple so that it is possible to give explicit formulas for $E(k, q, n)$. These we proceed to develop. Some of these were given by van der Corput and Kluyver **(4)**.

4. The case $k = 3$. By (6) and Theorem 1 we see that

$$E(3, 2, n) = E(3, 0, n), \qquad E(3, 1, n) = -2E(3, 0, n).$$

Hence it remains to find $E(3, 0, n)$.

THEOREM 5. *Let n be a non-exceptional number with respect to 3, then*

$$E(3, 0, n) = \begin{cases} -\frac{1}{4}\lambda(n)\,\theta(n), & 3 \mid n, \\ -\frac{1}{2}\lambda(n)\,\theta(n), & \text{otherwise.} \end{cases}$$

Proof. By (11) we have

$$E(3, 0, n) = \sum_{\delta|n} r_3(\delta)\,\mu(n/\delta). \tag{13}$$

Now let $n = 3^\alpha n_1$ where, since n is not exceptional and >3, n_1 is a non-empty product of primes of the form $3x - 1$ and $\alpha = 0$ or 1. If δ_1 be any divisor of n_1 then

$$\lambda(\delta_1) \equiv r_3(\delta_1) \pmod 3$$

and since $3 - 2r_3(\delta_1)$ takes on the values $+1$ or -1 and is congruent to $r_3(\delta_1) \pmod 3$, we have

$$\lambda(\delta_1) = 3 - 2r_3(\delta_1).$$

Hence, by (13),

$$\begin{aligned} -2E(3, 0, n) &= \sum_{\delta \mid n} (-2r_3(\delta))\, \mu(n/\delta) = \mu(3^\alpha) \sum_{\delta_1 \mid n_1} (-2r_3(\delta_1))\, \mu(n_1/\delta_1) \\ &= \mu(3^\alpha) \sum_{\delta_1 \mid n_1} (\lambda(\delta_1) - 3)\, \mu(n_1/\delta_1) = \mu(3^\alpha) \sum_{\delta_1 \mid n_1} \lambda(\delta_1)\, \mu(n_1/\delta_1) \\ &= \mu(3^\alpha)\, \lambda(n_1)\, \theta(n_1) = \lambda(n)\, \theta(n_1), \end{aligned}$$

by (8). Now

$$\theta(n_1) = \begin{cases} \frac{1}{2}\theta(n), & \text{if } 3 \mid n, \\ \theta(n), & \text{otherwise.} \end{cases}$$

From this the theorem follows.

It follows from the above that $E(3, q, n)$ vanishes only for exceptional numbers n and then vanishes for all q.

5. The case $k = 4$. By use of (6) and Theorem 1 we find

$$E(4, 1, n) = -E(4, 0, n),\ E(4, 2, n) = -E(4, 0, n),\ E(4, 3, n) = E(4, 0, n).$$

Hence it suffices to consider $E(4, 0, n)$.

THEOREM 6. *Let $n > 4$ be a non-exceptional number with respect to 4. Then*

$$E(4, 0, n) = \begin{cases} -\lambda(n)\, \theta(n), & \textit{if } n \textit{ is odd}, \\ -\frac{1}{2}\lambda(n)\, \theta(n), & \textit{if } n \equiv 2 \pmod 4, \\ 0, & \textit{otherwise.} \end{cases}$$

Proof. Let $n = 2^\alpha n_1$ where n_1 is a product of primes of the form $4x - 1$. By (11)

$$E(4, 0, n) = \sum_{\delta \mid n} r_4(\delta)\, \mu(n/\delta).$$

Since $r_4(\delta)$ vanishes when δ is a multiple of 4 and is equal to 2 for other even δ, we may write, in view of (7)

$$E(4, 0, n) = \mu(2^\alpha) \sum_{\delta_1 \mid n_1} r_4(\delta_1)\, \mu(n_1/\delta_1).$$

Now, as in the proof of Theorem 5,

$$2 - r_4(\delta_1) = \lambda(\delta_1).$$

Hence, by (7) and (8),

$$E(4, 0, n) = -\mu(2^\alpha) \sum_{\delta_1 \mid n_1} (\lambda(\delta_1) - 2)\, \mu(n_1/\delta_1)$$
$$= -\mu(2^\alpha)\, \lambda(n_1)\, \theta(n_1)$$
$$= (-1)^{\alpha-1} \lambda(n)\, \mu(2^\alpha)\, \theta(n)\, 2^{-\alpha}.$$

Considering separately the cases $\alpha = 0$, $\alpha = 1$ and $\alpha > 1$, we have the results stated in the theorem.

6. The case $k = 6$. If we apply Theorem 1 we find

(14) $E(6, 5, n) = E(6, 0, n)$, $E(6, 4, n) = E(6, 1, n)$, $E(6, 3, n) = E(6, 2, n)$.

Since

(15) $$\phi(6, 0, n) + \phi(6, 1, n) = \phi(3, 0, n),$$

we have, by (4),

$$E(6, 0, n) + E(6, 1, n) = 2E(3, 0, n).$$

Hence by (6), (14), and (15), we have

$$E(6, 2, n) = -2E(3, 0, n),\ E(6, 1, n) = 2E(3, 0, n) - E(6, 0, n).$$

Thus it remains only to find $E(6, 0, n)$.

We have, as a special case of (11),

(16) $$E(6, 0, n) = \sum_{\delta \mid n} r_6(\delta)\, \mu(n/\delta).$$

Let us write $n = 2^\alpha 3^\beta m$ where m is prime to 6. We distinguish 5 cases.

Case I. $\alpha = 0, \beta = 0$. In this case the sum (16) extends over divisors δ which are prime to 6 so that

$$r_6(\delta) = 4r_3(\delta) - 3.$$

Since $n > 1$, we have

(17) $$E(6, 0, n) = 4 \sum_{\delta \mid n} r_3(\delta)\, \mu(n/\delta) = 4E(3, 0, n).$$

Case II. $\alpha = 1, \beta = 0$. In this case we note that if h is even

(18) $$r_6(h) = 2r_3(\tfrac{1}{2}h).$$

Hence

$$E(6, 0, n) = \sum_{\delta \mid n} \{2r_3(\delta) - r_6(\delta)\}\, \mu(n/(2\delta))$$
$$= 2E(3, 0, \tfrac{1}{2}n) - E(6, 0, \tfrac{1}{2}n).$$

Using case I we find

(19) $$E(6, 0, n) = -2E(3, 0, \tfrac{1}{2}n).$$

Case III. $\alpha \geq 2, \beta = 0$. In this case n/δ contains the factor 4 when δ is odd so that (16) has a contribution from only the even divisors δ. That is,

$$E(6, 0, n) = \sum_{\delta \mid \frac{1}{2}n} r_6(2\delta)\, \mu(n/(2\delta)) = 2E(3, 0, \tfrac{1}{2}n)$$

by (18).

Case IV. $\beta = 1$. Here we note that

$$r_6(3h) = 3r_2(h), \tag{20}$$

and write

$$\begin{aligned} E(6, 0, n) &= \sum_{\delta \mid \frac{1}{3}n} \{r_6(3\delta) - r_6(\delta)\}\, \mu(n/(3\delta)) \\ &= 3E(2, 0, \tfrac{1}{3}n) - E(6, 0, \tfrac{1}{3}n) \\ &= -E(6, 0, \tfrac{1}{3}n). \end{aligned}$$

Thus case IV reduces to one of the preceding cases.

Case V. $\beta \geq 2$. In this case n/δ contains the factor 9 when δ is not a multiple of 3, so that (16) becomes

$$E(6, 0, n) = \sum_{\delta \mid \frac{1}{3}n} r_6(3\delta)\, \mu(n/(3\delta)) = 3E(2, 0, \tfrac{1}{3}n) = 0,$$

in view of (20). Summing up the results of the 5 cases and applying Theorem 5 we have

THEOREM 7. *Let $n > 6$ be a non-exceptional number with respect to 6. Write $n = 2^\alpha 3^\beta n_1$ where n_1 is prime to 6. Then*

$$E(6, 0, n) = 2\mu^2(3^\beta)\, \frac{1 + 5\mu(2^\alpha)}{1 - 7\mu(2^\alpha)}\, \lambda(n)\, \theta(n_1).$$

We see that the non-zero values of $|E(6, q, n)|$ are powers of 2 as in the cases $k = 3, 4$.

7. Additional explicit formulas. Explicit formulas for $E(k, q, n)$ in case $\phi(k) > 2$ are in general lacking. We may remark, however, that

$$\begin{aligned} E(12, 2, n) &= 3E(4, 0, n) - 2E(6, 0, n) = E(12, 9, n) \\ E(12, 3, n) &= 4E(3, 0, n) - 3E(4, 0, n) = E(12, 8, n), \end{aligned}$$

so that $E(12, q, n)$ may be evaluated explicitly in the four cases $q = 2, 3, 8, 9$.

The case where n is a product of distinct primes of the form $kx - 1$ is however capable of treatment. In fact we have the following theorem.

THEOREM 8. *If n is the product of distinct primes of the form $kx - 1$ then:*

$$\begin{aligned} E(k, 0, n) &= E(k, k - 1, n) = \tfrac{1}{2}(2 - k)\, \mu(n)\, \theta(n) \\ E(k, 1, n) &= E(k, 2, n) = \ldots = E(k, k - 2, n) = \mu(n)\, \theta(n). \end{aligned}$$

Proof. Since every prime factor of n is of the form $kx - 1$, every divisor δ of n is of the form

$$\delta = m_\delta k + \mu(\delta), \tag{21}$$

where m_δ is an integer ≥ 0.

If $q = 0$ then (9) gives

$$\begin{aligned} E(k, 0, n) &= \sum_{\delta \mid n} \mu(n/\delta)\left\{\delta - k\left[\frac{\delta}{k}\right]\right\} \\ &= \sum_{\delta \mid n} \mu(n/\delta)\{\delta - k(m_\delta + \tfrac{1}{2}(\mu(\delta) - 1))\} \\ &= \sum_{\delta \mid n} \mu(n/\delta)\{\mu(\delta) - \tfrac{1}{2}k\mu(\delta)\} \\ &= \mu(n)\,\theta(n)(1 - \tfrac{1}{2}k). \end{aligned}$$

By Theorem 1, $E(k, k-1, n) = E(k, 0, n)$. If now $0 < q < k - 1$, we may show that

$$\left[\frac{q\delta}{k}\right] - \left[\frac{(q+1)\delta}{k}\right]$$

is not a function of q as follows. By (21)

$$\left[\frac{q\delta}{k}\right] = \left[qm_\delta + \frac{q\mu(\delta)}{k}\right] = qm_\delta + \tfrac{1}{2}(\mu(\delta) - 1),$$

$$\left[\frac{(q+1)\delta}{k}\right] = \left[qm_\delta + m_\delta + \frac{(q+1)\,\mu(\delta)}{k}\right] = qm_\delta + m_\delta + \tfrac{1}{2}(\mu(\delta) - 1).$$

Hence

$$\left[\frac{(q+1)\delta}{k}\right] - \left[\frac{q\delta}{k}\right] = m_\delta$$

is not a function of q. Therefore by (9)

$$E(k, 1, n) = E(k, 2, n) = \ldots = E(k, k-2, n).$$

To find this common value we need only use the fact that the sum of all the E's is zero. Thus

$$(k-2)\,E(k, 1, n) + 2E(k, 0, n) = 0,$$

that is,

$$(k-2)\,E(k, 1, n) = 2(\tfrac{1}{2}k - 1)\,\mu(n)\,\theta(n)$$

or

$$E(k, 1, n) = \mu(n)\,\theta(n).$$

Thus the proof is complete.

The explicit values obtained above for $E(k, 0, n)$ show that for an infinity of k and n, $|E| \neq 0$ and for these values

$$E(k, 0, n) \neq o(\theta(n)),$$

as $\theta(n) \to \infty$. This contradicts a conjecture of Erdős. Vijayaraghavan **(5)** showed the invalidity of the conjecture by a different argument in 1951.

8. General estimates. On the other hand we give some general results which show that totatives are, after all, fairly evenly distributed. Thus the following theorem shows that any two ϕ's do not differ by as much as $\theta(n) = O(n^{\epsilon})$ for any k.

THEOREM 9. $|\phi(k, q_1, n) - \phi(k, q_2, n)| \leqslant \theta(n)$.

Proof. Denote by $\{q_1, q_2\}$ the expression

$$[(q_1 + 1)\,\delta/k] - [q_1\,\delta/k] - [\delta/k] - [(q_2 + 1)\,\delta/k] + [q_2\,\delta/k] + [\delta/k].$$

For any real x, y the function

$$[x + y] - [x] - [y]$$

takes on the values 0 or 1. Hence $\{q_1, q_2\}$ takes on only 0, 1 or -1.

But by (10),

$$\phi(k, q_1, n) - \phi(k, q_2, n) = \sum_{\delta|n}{}' \mu(n/\delta)\{q_1, q_2\},$$

where the dash indicates that the summation extends over those divisors δ of n for which $\mu(n/\delta) \neq 0$. Hence

$$|\phi(k, q_1, n) - \phi(k, q_2, n)| \leqslant \sum_{\delta|n}{}' |\mu(n/\delta)||\{q_1, q_2\}| \leqslant \sum_{\delta|n}{}' 1 = \theta(n).$$

As a consequence of Theorem 9 we have

THEOREM 10. *For every* q, $|E(k, q, n)| \leqslant (k - 1)\,\theta(n)$.

Proof.

$$|E(k, q, n)| = |\phi(n) - k\phi(k, q, n)| = \left|\sum_{q_1=0}^{k-1} \{\phi(k, q_1, n) - \phi(k, q, n)\}\right|$$
$$\leqslant \sum_{\substack{q_1=0 \\ q_1 \neq q}}^{k-1} |\phi(k, q_1, n) - \phi(k, q, n)| \leqslant (k - 1)\,\theta(n).$$

As a corollary we have

$$\frac{\phi(n) - (k - 1)\,\theta(n)}{k} < \phi(k, q, n) < \frac{\phi(n) + (k - 1)\,\theta(n)}{k},$$

uniformly in q. For $q = 0$ we have the stronger statement,

$$\frac{\phi(n)}{k} - \tfrac{1}{2}\,\theta(n) < \phi(k, 0, n) < \frac{\phi(n)}{k} + \tfrac{1}{2}\,\theta(n). \tag{22}$$

In fact, by (10),

$$\phi(k, 0, n) = \sum_{\delta|n} \mu(n/\delta)\left[\frac{\delta}{k}\right]$$
$$= \frac{1}{k}\sum_{\delta|n} \delta\mu(n/\delta) - \sum_{\delta|n} \mu(n/\delta)\left\{\frac{\delta}{k} - \left[\frac{\delta}{k}\right]\right\} = \frac{\phi(n)}{k} - S.$$

Now

$$|S| < \sum_{\substack{\delta \mid n \\ \mu(n/\delta)=1}} \mu(n/\delta) = \tfrac{1}{2}\theta(n).$$

From this (22) follows at once.

9. Applications. We conclude with a few remarks about an application of the foregoing results.

Let $Q_n(x)$ denote the irreducible polynomial whose roots are the $\phi(n)$ primitive nth roots of unity. That is

$$Q_n(x) = \prod_{\tau} (x - \exp(2\pi i\tau/n))$$

or

$$Q_n(x) = \prod_{\tau<\frac{1}{2}n} (x^2 - 2x\cos(2\pi\tau/n) + 1). \tag{23}$$

We suppose that $n > 6$ and $x > 0$ and ask for inequalities for $Q_n(x)$. Since the factors of (23) are monotone increasing functions of τ, inequalities are easily obtained by subdividing the range $0 \leqslant \tau < \frac{1}{2}n$ into (for example) the four intervals

$$0 \leqslant \tau < \tfrac{1}{8}n,\ \tfrac{1}{8}n \leqslant \tau < \tfrac{1}{4}n,\ \tfrac{1}{4}n \leqslant \tau < \tfrac{3}{8}n,\ \tfrac{3}{8}n \leqslant \tau < \tfrac{1}{2}n$$

and counting the number of totatives in these intervals. These numbers are respectively:

$$\begin{aligned} A &= \phi(6, 0, n), \\ B &= \phi(4, 0, n) - \phi(6, 0, n), \\ C &= \phi(3, 0, n) - \phi(4, 0, n), \\ D &= \phi(6, 2, n) = \tfrac{1}{2}\phi(n) - \phi(3, 0, n). \end{aligned}$$

Thus we obtain the following inequalities:

$$Q_n(x) > (x-1)^{2A}(x^2 - x + 1)^B(x^2 + 1)^C(x^2 + x + 1)^D, \tag{24}$$

$$Q_n(x) < (x^2 - x + 1)^A(x^2 + 1)^B(x^2 + x + 1)^C(x + 1)^{2D}. \tag{25}$$

Estimates for A, B, C, D, may be obtained from (22) and give

$$\begin{aligned} \tfrac{1}{6}\phi(n) - \tfrac{1}{2}\theta(n) &\leqslant A \leqslant \tfrac{1}{6}\phi(n) + \tfrac{1}{2}\theta(n), \\ \tfrac{1}{12}\phi(n) - \theta(n) &\leqslant B \leqslant \tfrac{1}{12}\phi(n) + \theta(n), \\ \tfrac{1}{12}\phi(n) - \theta(n) &\leqslant C \leqslant \tfrac{1}{12}\phi(n) + \theta(n), \\ \tfrac{1}{6}\phi(n) - \tfrac{1}{2}\theta(n) &\leqslant D \leqslant \tfrac{1}{6}\phi(n) + \tfrac{1}{2}\theta(n). \end{aligned}$$

Sharper inequalities, especially for certain types of n, can be obtained from (24) and (25) by applying Theorems 2, 4, 5, 6, and 7. Similar results may be written down for $x < 0$. Such results are useful in discussing the existence of "characteristic prime factors" of $a^n - b^n$, Lucas's functions and their generalizations **(2)**. Of course any such inequalities will not give the asymptotically correct result:

$$Q_n(x) = x^{\phi(n)}(1 - \mu(n)x^{-1} + O(x^{-2})) \qquad (x \to \infty).$$

Their utility lies in the direction of actual inequalities for a fixed value of x.

References

1. A. M. Legendre, *Essai sur la théorie des nombres* (2nd ed., Paris, 1808).
2. D. H. Lehmer, *On the converse of Fermat's theorem* II, Amer. Math. Monthly, *56* (1949), 304–306.
3. J. Liouville, *Sur quelques fonctions numériques*, J. de Math (2), *2* (1857), 244–248.
4. J. G. van der Corput and J. C. Kluyver, *Vraagstuk* CXCII, Wiskunde Opgaven, *11* (1912–14), 483–488.
5. T. Vijayaraghavan, *On a problem in elementary number theory*, J. Indian Math. Soc., *15* (1951), 51–56.

University of California

SYSTEMS OF LINEAR CONGRUENCES

A. T. BUTSON AND B. M. STEWART

1. Introduction. On recent occasions papers have been presented concerned with the problem of solving a system of linear congruences. Apparently the authors were not aware that this problem was solved very neatly and completely a long time ago by H. J. S. Smith **(5; 6)**. One reason for this situation is that recent texts in the theory of numbers go only as far in the discussion of systems of congruences as one can with the most elementary tools; whereas older texts, such as the one by Stieltjes **(8, pp. 284–377)**, devote so much space to the discussion of the requisite matric theory that the reader is liable to lose sight of the elegant results concerning systems of congruences. Perhaps the time has come to give a new exposition of this material, particularly since this can be done in rather short compass to an audience whose background may be assumed to include acquaintance with the invariant factors and the Smith normal form of a matrix with elements in a principal ideal ring, $\mathfrak{P}$.

In the final part of this paper we present some original work extending the discussion of systems of linear equations, and systems of linear congruences modulo an ideal, from the classical case over the rational domain to the case where the systems are over a set of integral elements, with a $\mathfrak{P}$-basis, belonging to an associative algebra. Here we assume knowledge of the Hermite normal form of a matrix with elements in a principal ideal ring.

THE CLASSICAL CASE

2. The problems. The coefficients, constants, and moduli in the following equations and congruences are assumed to be in a specified principal ideal ring $\mathfrak{P}$, such as the rational domain. For the system of p linear equations in n unknowns represented by

$$\sum_{i=1}^{n} x_i a_{ij} = k_j, \qquad j = 1, 2, \ldots, p,$$

we will use the matric notation

$$XA = K, \tag{1}$$

where X is 1-by-n, A is n-by-p, and K is 1-by-p.

The first problem is to determine when (1) has a solution X with elements in $\mathfrak{P}$, to find how many solutions there may be, and to give a method for actually obtaining the solution.

For the system of linear congruences represented by

$$\sum_{i=1}^{n} x_i b_{ij} \equiv g_j \pmod{m_j}, \qquad j = 1, 2, \ldots, p,$$

Received November 15, 1954.

we note that if $m = [m_1, m_2, \ldots, m_p]$ is the least common multiple of the m_j, then an equivalent system of congruences is given by

$$\sum_{i=1}^{n} x_i a_{ij} \equiv k_j \pmod{m}, \qquad j = 1, 2, \ldots, p,$$

where $a_{ij} = b_{ij}r_j$, $k_j = g_j r_j$, and $m = m_j r_j$. If we denote this system by

$$XA \equiv K \pmod{m}, \tag{2}$$

then the second problem is to answer for (2) the same three questions we have listed for (1).

3. Necessary and sufficient conditions for solving (1). There exist (4, Theorem 105.2) unimodular matrices U and V with elements in $\mathfrak{P}$, U being n-by-n and V being p-by-p, such that $UAV = E$ is in Smith normal form, with zero elements everywhere except in the main diagonal where there may appear non-zero elements $e_1, e_2, \ldots, e_r$ (which are called invariant factors and which are uniquely determined up to associates in $\mathfrak{P}$) having the property that e_i divides e_{i+1} and either $r \leqslant p \leqslant n$ or $r \leqslant n < p$.

Hence the system (1) may be replaced by the equivalent system

$$(XU^{-1})(UAV) = KV,$$

so that by setting $Y = XU^{-1}$ and $C = KV$, we arrive at

$$YE = C. \tag{3}$$

The system (3) is so simple that we can immediately conclude that necessary and sufficient conditions for its solution are as follows:

$$e_i \text{ must divide } c_i, \quad i = 1, 2, \ldots, r; \quad c_i = 0, i > r. \tag{4}$$

If we define $A' = \binom{A}{K}$ as the augmented matrix of (1), then using the conventional block notation we have

$$\begin{pmatrix} U & O \\ O & 1 \end{pmatrix} A' V = \begin{pmatrix} E \\ C \end{pmatrix},$$

so that a further transformation by unimodular matrices U' and V', where U' is $(n+1)$-by-$(n+1)$ and V' is p-by-p, will take A' into its Smith normal form, say $U'\binom{E}{C}V' = E'$, which is $(n+1)$-by-p with elements e'_i in the main diagonal. Thus depending on the relative size of n and p, the conditions (4) may be given the following form:

(5) $p \leqslant n$: $e'_i = e_i$, $i = 1, 2, \ldots, p$;

(5′) $n < p$: $e'_i = e_i$, $i = 1, 2, \ldots, n$; and $e'_{n+1} = 0$.

4. Necessary and sufficient conditions for solving (2). Since the congruence problem (2) requires the existence of elements t_j in $\mathfrak{P}$ such that

$$t_j m + \sum_{i=1}^{n} x_i a_{ij} = k_j, \qquad j = 1, 2, \ldots, p,$$

it is easy to replace the system of congruences (2) by an equivalent system of $p_1 = p$ equations in $n_1 = n + p$ unknowns, say

$$XA + TM = K, \tag{6}$$

where T is 1-by-p and $M = mI_p$ is a scalar p-by-p matrix.

Since $p_1 < n_1$, we apply the test (5) to the system (6). This requires us to compute the invariant factors of $\binom{A}{M}$ and $\binom{A'}{M}$. Fortunately this task is easy because of the form of M. Following an argument by Butson which is more direct than that used by Smith, we write

$$\begin{pmatrix} U & 0 \\ 0 & V^{-1} \end{pmatrix} \begin{pmatrix} A \\ M \end{pmatrix} V = \begin{pmatrix} E \\ M \end{pmatrix}.$$

Because e_i divides e_{i+1}, we see that no further arrangement of columns is necessary and that the ith invariant factor of $\binom{A}{M}$ is either (e_i, m) when $i \leqslant p < n$, or is m when $n < i \leqslant p$.

Similarly, for $\binom{A'}{M}$ the ith invariant factor is (e'_i, m) when $i \leqslant p \leqslant n$; but when $n < p$, the $(n+1)$st invariant factor is (e'_{n+1}, m), and when $n + 1 < i \leqslant p$, the ith invariant factor is m.

Hence the test (5) shows that the necessary and sufficient conditions for the solution of (2) are as follows:

(7) $p \leqslant n$: $(e'_i, m) = (e_i, m)$, $i = 1, 2, \ldots, p$;

(7′) $n < p$: $(e'_i, m) = (e_i, m)$, $i = 1, 2, \ldots, n$; and $(e'_{n+1}, m) = m$.

We note that the final condition in (7′) may be written $e'_{n+1} \equiv 0 \pmod{m}$.

5. The number of solutions and their form. To determine how many solutions there are and actually to produce them, we return to (3), supposing that the necessary and sufficient conditions stated above are satisfied.

In the case of equations we see from $YE = C$, that the first r of the y's are determined uniquely by $y_i = c_i/e_i$, while the remaining $n - r$ of the y's are arbitrary. The complete solution of (1) is then given by $X = YU$ and involves $n - r$ parameters. Of course, since U and V are not unique, the complete solution may be obtained in a variety of forms, differently expressed, but actually equivalent.

In the case of congruences we consider solutions X' and X of (2) to be distinct only when $X' \not\equiv X \pmod{m}$, i.e., when for at least one value of i we have $x'_i \not\equiv x_i \pmod{m}$. We see from $YE \equiv C \pmod{m}$, that the first r of the y's are determined by congruences of the form $y_i e_i \equiv c_i \pmod{m}$. From properties of $\mathfrak{P}$ we know there are as many solutions y_i which are incongruent mod m as there are residue classes of $\mathfrak{P}$, mod (e_i, m). The remaining y's are arbitrary, so for each of these there are as many solutions incongruent mod m as there are residue classes of $\mathfrak{P} \pmod{m}$. The solutions of (2) are given explicitly by $X \equiv YU \pmod{m}$, so there are as many distinct solutions X as there are distinct solutions Y. (Moreover, we may check that $x'_i \equiv x_i \pmod{m}$ if and only if

$$x'_j \equiv x_j \pmod{m_j}, \qquad j = 1, 2, \ldots, p;$$

so there are the same number of incongruent solutions of the original system of congruences with moduli $m_1, m_2, \ldots, m_p$.)

In particular, when $\mathfrak{P}$ is the domain of rational integers, the above considerations show that there are exactly

$$N = (e_1, m)(e_2, m) \ldots (e_r, m)\, m^{n-r}$$

distinct solutions of (2).

6. Example. We take $\mathfrak{P}$ to be the rational integers and consider the system

$$3x_1 + x_2 = 5,$$
$$5x_1 + 3x_2 = 1.$$

Using the notation of the preceding sections, we have

$$UAV = \begin{pmatrix} 0 & 1 \\ 1 & -3 \end{pmatrix}\begin{pmatrix} 3 & 5 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 3 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} = E,$$

$$KV = (5 \quad 1)\, V = (5 \quad 14) = C,$$

$$U'\begin{pmatrix} E \\ C \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -5 & -3 & 1 \\ 10 & 7 & -2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 4 \\ 5 & 14 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & 0 \end{pmatrix} = E'.$$

Since $4 = e_2 \neq e'_2 = 2$, it follows from (5) that the system has no solution in rational integers.

Considering the same system mod m, we see from (7) that we must have $(1, m) = (1, m)$ and $(4, m) = (2, m)$. If $m \equiv 0 \pmod 4$, there are no solutions; if $m \equiv 1$ or $3 \pmod 4$, there will be $N = 1$ solutions; and if $m \equiv 2 \pmod 4$, there will be $N = 2$ solutions.

Thus if $m = 10$, we solve $y_1 \equiv 5$, $4y_2 \equiv 14$, for $y_1 \equiv 5$, $y_2 \equiv 1$; and $y_1 \equiv 5$, $y_2 \equiv 6$. Then from $X = YU$, we compute $x_1 \equiv 1$, $x_2 \equiv 2$; and $x_1 \equiv 6$, $x_2 \equiv 7$, respectively.

The Case of Integral Elements of an Algebra

7. Sets of integral elements. Let $\mathfrak{A}$ be an associative algebra defined over a field $\mathfrak{F}$ and possessing a modulus ϵ. Suppose that $\mathfrak{F}$ contains the principal ideal ring $\mathfrak{P}$. Each element of a set $\mathfrak{Q}$ of elements of $\mathfrak{A}$ will be called an integral element if the set has the following three properties:

U (unity): the set contains the modulus ϵ;

C (closure): the set is closed under addition, subtraction, and multiplication;

B (finite basis): the set contains elements $\epsilon_1 = \epsilon, \epsilon_2, \ldots, \epsilon_k$ such that every element of the set is expressed uniquely in the form

$$\sum a_i \epsilon_i$$

where each a_i is in $\mathfrak{P}$.

As an example we may take $k = 1$ and obtain as $\mathfrak{Q}$ the ring $\mathfrak{P}$ itself. Again when $\mathfrak{F}$ is the rational field and $\mathfrak{P}$ the rational domain, we see that $\mathfrak{Q}$ is a set

of integral elements (but not necessarily a maximal set) in the sense of Dickson (1, Chapter X). See also (3).

8. The regular representations. If $\alpha = \sum a_i \epsilon_i$ is any element of $\mathfrak{Q}$, by properties C and B there must exist elements r_{ij} and s_{ji} of $\mathfrak{P}$ such that

$$\alpha \epsilon_j = (\sum a_i \epsilon_i) \epsilon_j = \sum r_{ij} \epsilon_i, \qquad j = 1, 2, \ldots, k;$$
$$\epsilon_j \alpha = \epsilon_j (\sum a_i \epsilon_i) = \sum s_{ji} \epsilon_i, \qquad j = 1, 2, \ldots, k.$$

Hence with each α in $\mathfrak{Q}$ there are associated k-by-k matrices $R(\alpha) = (r_{ij})$ and $S(\alpha) = (s_{ji})$ with elements in $\mathfrak{P}$.

If $\mathbf{E}$ indicates the 1-by-k matrix with elements $\epsilon_1, \epsilon_2, \ldots, \epsilon_k$, then in matric notation we have

$$\alpha \mathbf{E} = \mathbf{E} R(\alpha), \quad \mathbf{E}^T \alpha = S(\alpha) \mathbf{E}^T. \tag{8}$$

Since $\epsilon_1 = \epsilon$, the first column of $R(\alpha)$ and the first row of $S(\alpha)$ consist of precisely the elements $a_1, a_2, \ldots, a_k$. Then from property B it follows that the correspondences defined by (8) are both one-to-one. Moreover, it is easily shown that the correspondences are preserved under the addition and multiplication operations of each system. Hence the matrices $R(\alpha)$ and the matrices $S(\alpha)$ provide isomorphic representations for $\mathfrak{Q}$, well known, respectively, as the first and second regular representations.

If α and β are in $\mathfrak{Q}$, we may use (8) and the fact that elements of $\mathfrak{P}$ commute with elements of $\mathfrak{Q}$ to write

$$R^T(\alpha) S(\beta) \mathbf{E}^T = R^T(\alpha) \mathbf{E}^T \beta = \alpha I \mathbf{E}^T \beta$$
$$= \alpha I S(\beta) \mathbf{E}^T = S(\beta) \alpha I \mathbf{E}^T = S(\beta) R^T(\alpha) \mathbf{E}^T;$$

then from property B, it follows that

$$R^T(\alpha) S(\beta) = S(\beta) R^T(\alpha).$$

In particular, letting $A = (a_1, a_2, \ldots, a_k)$ and $B = (b_1, b_2, \ldots, b_k)$ be the first rows of $R^T(\alpha)$ and $S(\beta)$, respectively, we obtain the useful relation

$$AS(\beta) = BR^T(\alpha). \tag{9}$$

9. Systems of linear equations over $\mathfrak{Q}$. We consider the following system of p linear equations in n unknowns:

$$\sum_{i=1}^{n} \alpha_{ij} \chi_i \beta_{ij} = \gamma_j, \qquad j = 1, 2, \ldots, p, \tag{10}$$

where the α_{ij}, β_{ij}, and γ_j are given elements of $\mathfrak{Q}$. Since $\mathfrak{Q}$ is not necessarily commutative, note that coefficients are allowed on both sides of the unknowns. We are concerned to establish necessary and sufficient conditions that (10) have solutions $\chi_1, \chi_2, \ldots, \chi_n$ which are in $\mathfrak{Q}$.

If we assume that such solutions exist we may write $\chi_i = \sum x_{ij} \epsilon_j$, where the x_{ij} are in $\mathfrak{P}$, and define $X_i = (x_{i1}, \ldots, x_{ik})$. Supposing $\gamma_j = \sum c_{ji} \epsilon_i$, we

define $C_j = (c_{j1}, \ldots, c_{jk})$. We define A_{ij} to be the first row of $S(\alpha_{ij})$. Then (8), (9) and (10) imply that

$$C_j \mathbf{E}^\tau = \sum A_{ij} \mathbf{E}^\tau \chi_i \beta_{ij} = \sum A_{ij} S(\chi_i) \mathbf{E}^\tau \beta_{ij}$$
$$= \sum X_i R^\tau(\alpha_{ij}) S(\beta_{ij}) \mathbf{E}^\tau.$$

Hence property B implies that

$$C_j = \sum_{i=1}^{n} X_i R^\tau(\alpha_{ij}) S(\beta_{ij}), \qquad j = 1, 2, \ldots, p.$$

We set

$$C = (c_{11}, \ldots, c_{1k}; c_{21}, \ldots, c_{2k}; c_{p1}, \ldots, c_{pk}),$$
$$X = (x_{11}, \ldots, x_{1k}; x_{21}, \ldots, x_{2k}; \ldots; x_{n1}, \ldots, x_{nk}),$$

and $A = (R^\tau(\alpha_{ij}) S(\beta_{ij}))$ where C is 1-by-pk, X is 1-by-nk, and the "enlarged coefficient matrix" A is nk-by-pk made up of k-by-k blocks of which the one in the ij-position is $R^\tau(\alpha_{ij}) S(\beta_{ij})$. Then the equations obtained above may be written as the single matric equation

$$XA = C. \tag{11}$$

Except for the size of the matrices involved, (11) is precisely a system of the classical type (1) with kp equations in kn unknowns, with the elements involved all in $\mathfrak{P}$.

Conversely, if (11) has a solution X in $\mathfrak{P}$, we can retrace the steps above to obtain in $\mathfrak{Q}$ a solution of (10). Moreover, by property B, distinct solutions of (11) lead to distinct solutions of (10).

Thus the problem of solving (10) in $\mathfrak{Q}$ has been shown equivalent to solving (11) in $\mathfrak{P}$. Referring to (5) and (5′) we can assert that if $e_1, e_2, \ldots$ are the invariant factors of A and if $e'_1, e'_2, \ldots$ are the invariant factors of the augmented matrix $\binom{A}{C}$, then necessary and sufficient conditions that the system (10) have a solution are that

(12) $p \leqslant n$: $e'_i = e_i$, $i = 1, 2, \ldots, kp$;
(12′) $n < p$: $e'_i = e_i$, $i = 1, 2, \ldots, kn$; and $e_{kn+1}' = 0$.

Determining the number of solutions and the most general solution proceeds along the lines given in §5. In these matters it is worth a word of caution that the rank of A need not be a multiple of k.

We note, thanks to the referee, that one type of matric equation, well known in the literature, is included in the above discussion. For if the algebra $\mathfrak{A}$ is a total matric algebra of order $k = n^2$, having the natural basis of elements ϵ_{ij}, where ϵ_{ij} is an n-by-n matrix with 1 in the ij position and zeros elsewhere, so that the multiplication table is

$$\epsilon_{ij} \epsilon_{rs} = \delta_{jr} \epsilon_{is},$$

and if $\mathbf{E} = (\epsilon_{11}, \ldots, \epsilon_{1n}; \epsilon_{21}, \ldots, \epsilon_{2n}; \ldots; \epsilon_{n1}, \ldots, \epsilon_{nn})$, then the typical element $\beta = \sum b_{ij} \epsilon_{ij}$, which we ordinarily represent as $B = (b_{ij})$, has the regular representations

$$R(\beta) = I \cdot \times B, \quad S(\beta) = B \cdot \times I,$$

where $M \cdot \times B$ indicates the direct product matrix which is n^2-by-n^2 and whose ij block is Mb_{ij}. Then a linear equation like $\alpha \chi \beta = \gamma$ is replaced, according to the theory above for passing from (10) to (11), by an equation $X' D' = C'$, where

$$D' = R^T(\alpha)\, S(\beta) = (I \cdot \times A)^T (B \cdot \times I) = B \cdot \times A^T$$

and X' and C' are 1-by-n^2, obtained from $X = (x_{ij})$ and $C = (c_{ij})$, respectively, by taking row blocks. The general linear equation in one unknown $\sum \alpha_i \chi \beta_i = \gamma$ may be treated in the same manner, the enlarged coefficient matrix being $D'' = \sum B_i \cdot \times A_i^T$. This is the *nivellateur* studied by Sylvester **(9)**, however, only for the case $\mathfrak{P} = \mathfrak{F}$.

Similarly, the system of equations (10) may be generalized to allow each unknown to appear in a finite number of summands in each equation; the technique for passing to (11) remains the same, except each component block of the enlarged coefficient matrix will now be a sum of matrices of the type $R^T(\alpha_{ij})\, S(\beta_{ij})$.

10. Minimal bases for ideals in $\mathfrak{Q}$. In the usual manner the left ideal $\mathfrak{M}$ generated by $\zeta_1, \zeta_2, \ldots, \zeta_t$, a given set of elements of $\mathfrak{Q}$, is defined to be the set of elements

$$\sum_{i=1}^{t} \nu_i \zeta_i$$

obtained by allowing the left-multipliers ν_i to vary independently over all of $\mathfrak{Q}$. A minimal basis for the ideal $\mathfrak{M}$ is by definition a set of elements $\mu_1, \mu_2, \ldots, \mu_s$ such that an element of $\mathfrak{Q}$ is in the ideal $\mathfrak{M}$ if and only if it can be represented in the form

$$\sum_{i=1}^{s} c_i \mu_i,$$

where the c_i are in $\mathfrak{P}$; and this representation is to be unique.

An argument by MacDuffee **(2)** shows that if H is the uniquely determined left-Hermite form of the matrix

$$S = \begin{pmatrix} S(\zeta_1) \\ \cdots \\ S(\zeta_t) \end{pmatrix},$$

then the non-zero rows H_i of H determine a minimal basis for $\mathfrak{M}$, having $s \leqslant k$, by the relation $\mu_i = H_i \mathbf{E}^T$. The notation which we have been using makes it simple to reproduce the proof.

Let U be a unimodular matrix, kt-by-kt, having elements in $\mathfrak{P}$, such that $US = H$. Let $V = U^{-1}$, so that $S = VH$. If the ith row of U is divided into 1-by-k blocks U_{ij}, then

$$\mu_i = H_i \mathbf{E}^T = \sum U_{ij} S(\zeta_j)\, \mathbf{E}^T = \sum U_{ij} \mathbf{E}^T \zeta_j = \sum \nu_{ij} \zeta_j,$$

where $\nu_{ij} = U_{ij} \mathbf{E}^T$ is in $\mathfrak{Q}$, hence μ_i is in the ideal $\mathfrak{M}$, and so are all $\sum c_i \mu_i$.

Conversely, given any element $\nu = \sum \nu_i \zeta_i$ in the ideal, we have $\nu_i = \sum \pi_{ij} \epsilon_j$ and if we define $N_i = (\pi_{i1}, \ldots, \pi_{ik})$ we can write $\nu_i = N_i \mathbf{E}^T$. Hence

$$\nu = \sum N_i \mathbf{E}^T \zeta_i = \sum N_i S(\zeta_i) \mathbf{E}^T = NS\mathbf{E}^T = NVH\mathbf{E}^T = \sum c_i \mu_i$$

where N is 1-by-kt, made up of the 1-by-k blocks N_i, and where c_i is the element in the ith column of NV. Since c_i is in $\mathfrak{P}$, a representation of the desired type for ν has been found. The uniqueness of the representation follows from the independence of the non-zero rows in the canonical left-Hermite form H.

In an analogous way we define the right-ideal generated by $\zeta_1, \zeta_2, \ldots, \zeta_t$ to be the set of elements $\sum \zeta_i \eta_i$ obtained by allowing the right-multipliers η_i to vary independently over $\mathfrak{Q}$. In this case a minimal basis can be found by computing the left-Hermite form D of the matrix R which is kt-by-k with its ith k-by-k block being $R^T(\zeta_i)$; for if $D_1, \ldots, D_r$ are the non-zero rows of D, necessarily with $r \leqslant k$, then the elements $\delta_j = D_j \mathbf{E}^T$ serve as a minimal basis.

By combining these observations we can find a minimal basis for the two-sided ideal generated by $\zeta_1, \zeta_2, \ldots, \zeta_t$ whose typical element is

$$\alpha = \sum_{i=1}^{t} \sum_{j=1}^{q_i} \nu_{ij} \zeta_i \eta_{ij},$$

where the q_i are all finite. For we may first compute a minimal basis $\mu_1, \mu_2, \ldots, \mu_s$ for the left ideal generated by $\zeta_1, \zeta_2, \ldots, \zeta_t$ and replace each $\nu_{ij} \zeta_i$ by $\sum c_{ijm} \mu_m$. Then

$$\alpha = \sum_{m=1}^{s} \mu_m \eta_m, \text{ where } \eta_m = \sum_{i=1}^{t} \sum_{j=1}^{q_i} c_{ijm} \eta_{ij}.$$

Hence if, secondly, we compute a minimal basis $\delta_1, \delta_2, \ldots, \delta_r$ for the right ideal generated by $\mu_1, \mu_2, \ldots, \mu_s$, we shall have arrived at a suitable minimal basis $\delta_1, \delta_2, \ldots, \delta_r$ for the two-sided ideal generated by $\zeta_1, \zeta_2, \ldots, \zeta_t$.

However, not every matrix H in left-Hermite form represents a minimal basis for an ideal of $\mathfrak{Q}$ **(2, p. 76)**.

When α and β are in $\mathfrak{Q}$, by the notation $\alpha \equiv \beta \pmod{\mathfrak{M}}$ we mean that $\alpha - \beta$ is in the ideal $\mathfrak{M}$ and we say that α and β are in the same residue class mod $\mathfrak{M}$. For the sequel it is important to notice that, in general, it is only when the ideal $\mathfrak{M}$ is two-sided that multiplication of residue classes mod $\mathfrak{M}$ is well defined.

11. Systems of linear congruences modulo ideals. Over $\mathfrak{Q}$ we consider the following system of p linear congruences, modulo ideals of $\mathfrak{Q}$, in n unknowns:

$$\sum_{i=1}^{n} \alpha_{ij} \chi_i \beta_{ij} \equiv \gamma_j \pmod{\mathfrak{M}_j}, \qquad j = 1, 2, \ldots, p. \tag{13}$$

We shall assume as explained in §10 that for the ideal $\mathfrak{M}_j$, whether it be left, right, or two-sided, a minimal basis of s_j elements has been found, say

$$\mu_{1j}, \mu_{2j}, \ldots, \mu_{s_j j},$$

given by $\mu_{ij} = H_{ij} \mathbf{E}^T$ where the H_{ij} are non-zero rows of a left-Hermite

k-by-k matrix. We let H_j be the s_j-by-k matrix with rows H_{ij}, so that H_j is what is called an echelon row form.

The system (13) is then equivalent to the system

$$\sum_{i=1}^{n} \alpha_{ij}\, \chi_i\, \beta_{ij} + \sum_{i=1}^{s_j} t_{ij}\, \mu_{ij} = \gamma_j, \qquad j = 1, 2, \ldots, p,$$

where the unknowns t_{ij} are in $\mathfrak{P}$. Following the same development and using the same notation as in §9, we may show that solving (13) in $\mathfrak{Q}$ is equivalent to solving in $\mathfrak{P}$ the following system:

$$(X\ T)\begin{pmatrix} A \\ H \end{pmatrix} = C, \tag{14}$$

where

$$T = (t_{11}, \ldots, t_{s_1,1}; t_{12}, \ldots, t_{s_2,2}; \ldots; t_{1p}, \ldots, t_{s_p,p})$$

and H is the direct sum $H = H_1 \dotplus H_2 \dotplus \ldots \dotplus H_p$.

The number of unknowns in (14) is $nk + s$, where $s = \sum s_j$, and the number of equations is pk.

If $pk < nk + s$, we apply (5) to obtain the conditions

$$e_i\begin{pmatrix} A \\ C \\ H \end{pmatrix} = e_i\begin{pmatrix} A \\ H \end{pmatrix}, \qquad i = 1, 2, \ldots, pk. \tag{15}$$

If $nk + s < pk$, we apply (5′) to obtain the conditions

$$e_i\begin{pmatrix} A \\ C \\ H \end{pmatrix} = \begin{cases} e_i\begin{pmatrix} A \\ H \end{pmatrix}, & i = 1, 2, \ldots, nk + s, \\ 0, & i = nk + s + 1. \end{cases} \tag{15′}$$

Thus (15) and (15′) represent necessary and sufficient conditions for the solution of (13).

If the solution is obtained as in §5, starting from (14), unnecessary parameters may be noticed. We have made a further study of (14) in which the Smith forms D_j of H_j play a part, as well as the least common multiple m of the elements of all the D_j. This method seems of some interest because of avoiding unnecessary parameters, but the alternative set of conditions which is obtained lacks the directness of (15) and (15′). This further study emphasized the need of care in the definitions of congruent solutions.

Suppose that $\chi_1, \chi_2, \ldots, \chi_n$ is a solution of (13). If all the ideals $\mathfrak{M}_j$ are two-sided and if

$$\chi'_i \equiv \chi_i \pmod{\mathfrak{M}_j}, \qquad i = 1, 2, \ldots, n; j = 1, 2, \ldots, p; \tag{16}$$

then $\chi'_1, \chi'_2, \ldots, \chi'_n$ is also a solution of (16). But if one or more of the ideals $\mathfrak{M}_j$ is one-sided, (16) is no longer sufficient to guarantee that $\chi'_1, \chi'_2, \ldots, \chi'_n$ is a solution of (13). Having given these words of caution, we now define sets of solutions of (13) which satisfy (16) to be congruent sets.

In matric form (16) may be written

$$X'_i - X_i = W_{ij} H_j, \qquad i = 1, 2, \ldots, n; j = 1, 2, \ldots, p.$$

Hence $X'_i - X_i$ is a common left multiple of the H_j. By repeated application of the method described in **(7)**, there is a constructive way of finding M the least common left multiple of $H_1, H_2, \ldots, H_p$. Then (16) is equivalent to the existence of Q_i with elements in $\mathfrak{P}$ such that

$$X'_i - X_i = Q_i M, \qquad i = 1, 2, \ldots, n;$$

and hence to the single matric equation

$$X' - X = QM^* \tag{17}$$

where $M^* = M \dotplus M \dotplus \ldots \dotplus M$ with n summands. Hence we may apply (5) and (5′) to obtain conditions equivalent to (16) expressed in terms of the invariant factors of M^* and

$$\begin{pmatrix} M^* \\ X' - X \end{pmatrix}.$$

12. Example. Letting $\mathfrak{F}$ be the rational field and $\mathfrak{P}$ the rational domain, we consider the algebra $\mathfrak{A}$ having as a basis $\epsilon_1 = \epsilon, \epsilon_2, \epsilon_3$ with $\epsilon_2 \epsilon_2 = \epsilon_2$, $\epsilon_3 \epsilon_2 = \epsilon_3$, and $\epsilon_2 \epsilon_3 = \epsilon_3 \epsilon_3 = 0$. If we take as $\mathfrak{O}$ the set of all $\alpha = a \epsilon_1 + b \epsilon_2 + c \epsilon_3$ where a, b, c are in $\mathfrak{P}$, we have a set of integral elements with the basis $\epsilon_1, \epsilon_2, \epsilon_3$. Using (8) we find

$$R^r(\alpha) = \begin{pmatrix} a & b & c \\ 0 & a+b & c \\ 0 & 0 & a \end{pmatrix}, \qquad S(\alpha) = \begin{pmatrix} a & b & c \\ 0 & a+b & 0 \\ 0 & 0 & a+b \end{pmatrix}.$$

Taking $\alpha = (3, 3, 1)\,\mathbf{E}^r$, $\beta = (1, 5, 2)\,\mathbf{E}^r$, $\gamma = (0, 0, 2)\,\mathbf{E}^r$, $\zeta = (6, 2, 12)\,\mathbf{E}^r$, we will study

$$\alpha \chi \beta = \gamma, \tag{18}$$

$$\alpha \chi \beta \equiv \gamma \pmod{(\zeta]}, \tag{19}$$

$$\alpha \chi \beta \equiv \gamma \pmod{[\zeta)}, \tag{20}$$

where $(\zeta]$ and $[\zeta)$ indicate, respectively, the left and right ideals generated by ζ. First we compute

$$A = R^r(\alpha) S(\beta) = \begin{pmatrix} 3 & 3 & 1 \\ 0 & 6 & 1 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 5 & 2 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{pmatrix} = \begin{pmatrix} 3 & 33 & 12 \\ 0 & 36 & 6 \\ 0 & 0 & 18 \end{pmatrix}, \quad E = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 108 \end{pmatrix}.$$

When we find that $A' = \binom{A}{C}$ has $e'_1 = 1$, $e'_2 = 6$, $e'_3 = 108$, it follows from (12) that there is no solution to (18).

Since $S(\zeta)$ has the left-Hermite form

$$L = \begin{pmatrix} 24 & 0 & 0 \\ 12 & 4 & 0 \\ 6 & 2 & 4 \end{pmatrix},$$

we find that $\binom{A}{L}$ has $e_1 = 1, e_2 = 2, e_3 = 12$, and that $\binom{A'}{L}$ has $e'_1 = 1, e'_2 = 2, e'_3 = 12$, so by (15) there is a solution of (19).

Since $R^T(\xi)$ has the left-Hermite form

$$R = \begin{pmatrix} 24 & 0 & 0 \\ 6 & 2 & 0 \\ 0 & 0 & 6 \end{pmatrix},$$

we find that $\binom{A}{R}$ has $e_1 = 1, e_2 = 6, e_3 = 12$, but that $\binom{A'}{R}$ has $e'_1 = 1, e'_2 = 2, e'_3 = 12$, so by (15) there is no solution to (20).

When we carry through the actual solution of (19) we find that the most general solution involves three parameters:

$$x_1 = -24z_1 + 2z_3, \quad x_2 = 22z_1 - 2z_3, \quad x_3 = 1 + 78z_1 + 4z_2 - 12z_3.$$

When we apply (17) we find that pairs of the solutions are congruent mod $(\xi]$, if and only if

$$z'_1 \equiv z_1 (\text{mod } 4), \quad z'_2 \equiv z_2 (\text{mod } 2), \quad z'_3 \equiv z_3 (\text{mod } 3).$$

References

1. L. E. Dickson, *Algebras and their arithmetics*, (Chicago, 1923).
2. C. C. MacDuffee, *An introduction to the theory of ideals in linear associative algebras*, Trans. Amer. Math. Soc., *31* (1929), 71–90.
3. ———, *Modules and ideals in a Frobenius algebra*, Monatsh. Math., *48* (1939), 293–313.
4. ———, *An introduction to abstract algebra* (New York, 1940).
5. H. J. S. Smith, *On systems of linear indeterminate equations and congruences*, Phil. Trans. Royal Soc. London, A *151* (1861), 293–326.
6. ———, *On the arithmetical invariants of a rectangular matrix of which the constituents are integral numbers*, Proc. London Math. Soc., *4* (1873), 236–249.
7. B. M. Stewart, *A note on least common left multiples*, Bull. Amer. Math. Soc., *55* (1949), 587–591.
8. T. J. Stieltjes, *Oeuvres complètes*, II (Math. Soc. Amsterdam, 1918).
9. J. J. Sylvester, C.R. Acad. Sci., Paris, *99* (1884), 117–118, 409–412, 432–436, 527–529.

Michigan State College

SOME SERIES OF PARTIALLY BALANCED INCOMPLETE BLOCK DESIGNS

D. A. SPROTT

1. Introduction. The use of incomplete block designs for estimating and judging the significance of the difference of treatment effects is now a standard statistical technique. A special kind of incomplete block design is the Partially Balanced Incomplete Block Design (PBIBD) introduced in **(3)**. A PBIBD is an incomplete block design that obeys the following conditions:

(1) there are b blocks of k distinct varieties each;

(2) there are v varieties, each replicated r times;

(3) given any variety, the remaining ones fall into sets of n_i varieties each ($i = 1, 2, \ldots, m$) such that every variety of the ith set occurs λ_i times with the given variety, λ_i being independent of the given initial variety;

(4) given any two varieties which are ith associates (that is, occur together λ_i times), the number of varieties which are jth associates of the one and kth associates of the other is

$$p_{jk}^{i} = p_{kj}^{i}$$

and is independent of the original pair of varieties.

From the definition of a PBIBD, certain fundamental identities concerning the parameters of the design follow; for ease of reference, these are listed as

$$bk = rv, \tag{1.1}$$

$$v - 1 = \sum_{i=1}^{m} n_i, \tag{1.2}$$

$$r(k-1) = \sum_{i=1}^{m} \lambda_i n_i, \tag{1.3}$$

$$n_i - 1 = \sum_{k=1}^{m} p_{ik}^{i}, \tag{1.4}$$

$$n_j = \sum_{k=1}^{m} p_{jk}^{i}, \tag{1.5}$$

$$n_i p_{jk}^{i} = n_j p_{ik}^{j}. \tag{1.6}$$

In **(3)**, a module theorem was proved by which it is possible in certain cases to construct the entire PBIBD by adding elements of a module to a given initial block. The purpose of the present paper is to generalize the module theorem of **(3)** to the case $v \neq b$, and then to apply the methods of **(6)** to obtain some general series of PBIBD's.

Received August 15, 1954.

2. The general module theorem.

THEOREM 2.1. *Let M be an additive abelian group (module) consisting of n elements x_i and suppose that there exist m blocks $B_1, B_2, \ldots, B_m$ such that:*

(1) *every block contains k distinct elements;*

(2) *among the $mk(k-1)$ differences arising from the m blocks, just n_i of the non-zero elements of M are repeated λ_i times; we may denote these n_i elements by*

$$\alpha_1^{\,i}, \alpha_2^{\,i}, \ldots, \alpha_{n_i}^{\,i};$$

(3) *among the $n_i(n_i-1)$ differences*

$$\alpha_u^{\,i} - \alpha_w^{\,i} \qquad (u, w = 1, 2, \ldots, n_i; u \neq w)$$

every element of the set

$$\alpha_1^{\,e}, \alpha_2^{\,e}, \ldots, \alpha_{n_e}^{\,e}$$

occurs $p_{ii}^{\,e}$ times, while among the $n_i n_j$ differences

$$\alpha_u^{\,i} - \alpha_w^{\,j} \quad (u = 1, 2, \ldots, n_i; w = 1, 2, \ldots, n_j),$$

every element of the set

$$\alpha_1^{\,e}, \alpha_2^{\,e}, \ldots, \alpha_{n_e}^{\,e}$$

occurs $p_{ij}^{\,e}$ times.
Then we can form blocks

$$B_{g\theta} \qquad (g = 1, 2, \ldots, m; \theta \in M)$$

of elements x', where $x' = x + \theta$ (x ranging over the elements of B_g). Since there are m initial blocks and θ assumes n values, we thus get mn blocks $B_{g\theta}$ ($B_{g0} = B_g$), and these blocks form a PBIBD with parameters

$$v = n, b = mn, r = mk, k; n_i, \lambda_i; P_i = (p_{jk}^{\,i}) \qquad (i = 1, 2, \ldots, m).$$

Proof. The proof of Theorem 2.1 is omitted, since it is an exact parallel of the proof given in **(3)** for the special case in which $m = 1, v = b$.

In the series which we shall derive from Theorem 2.1, it will be convenient to reserve the letter p to denote a prime; M will always be a Galois Field $GF(p^e)$, and x will always denote a primitive element of M.

3. Series with $4m(4\lambda \pm 1) + 1 = p^e$.

THEOREM 3.1. *If $v = 4m(4\lambda + 1) + 1 = p^e$, and if among the 2λ expressions*

$$x^{4ms} - 1 = x^{e_s} \qquad (s = 1, 2, \ldots, 2\lambda)$$

there are $\lambda + a$ even and $\lambda - a$ odd powers of x, then the design with parameters

$$v = 4m(4\lambda + 1) + 1, \quad b = mv, \quad r = m(4\lambda + 1), \quad k = 4\lambda + 1;$$
$$n_1 = n_2 = 2m(4\lambda + 1), \quad \lambda_1 = \lambda + a, \quad \lambda_2 = \lambda - a;$$
$$P_1 = \begin{pmatrix} \frac{1}{4}(v-5) & \frac{1}{4}(v-1) \\ \frac{1}{4}(v-1) & \frac{1}{4}(v-1) \end{pmatrix}, \quad P_2 = \begin{pmatrix} \frac{1}{4}(v-1) & \frac{1}{4}(v-1) \\ \frac{1}{4}(v-1) & \frac{1}{4}(v-5) \end{pmatrix}$$

can be constructed from the initial blocks

$$(x^{2i}, x^{2i+4m}, \ldots, x^{2i+16\lambda m})$$

where i ranges from 0 to $m-1$.

Proof. As in (6, Theorem 5.1), we find, setting $d = 4\lambda - s + 1$, that

$$x^{qd} = x^{q_s+2m(d-s)}.$$

Divide the elements of GF(v) into two sets; set 1 will be composed of elements of the form x^{2i}, and set 2 will be composed of elements of the form x^{2i+1}. The differences

$$x^{2i+4mr+q_s}, x^{2i+2m(4\lambda-2s+1+2r)+q_s},$$

for s fixed, range over set 1 or set 2 according as q_s is even or odd. Hence, as s ranges, each element of set 1 occurs $\lambda_1 = \lambda + a$ times and each element of set 2 occurs $\lambda_2 = \lambda - a$ times. Also, the number of elements in set 1 = the number of elements in set 2 $= \frac{1}{2}(v-1) = 2m(4\lambda+1)$. Thus condition 2 of Theorem 2.1 is satisfied and $n_1 = n_2 = 2m(4\lambda+1)$.

The differences between elements of set 1 have the form

$$x^{2i+2u} - x^{2i} = x^{2i}(x^{2u}-1) \quad (i = 0, 1, \ldots, \tfrac{1}{2}(v-3); u = 1, 2, \ldots, \tfrac{1}{2}(v-3)).$$

As i ranges, these differences cover set 1 or set 2 according as $x^{2u}-1$ is an even or an odd power of x. Hence, if y of the expressions $x^{2u}-1$ are odd powers of x, we have

$$p_{11}^{\ 2} = y, \quad p_{11}^{\ 1} = \tfrac{1}{2}(v-3) - y.$$

The differences between elements of set 2 have the form

$$x^{2i+2u+1} - x^{2i+1} = x^{2i+1}(x^{2u}-1).$$

These cover set 1 or set 2 according as $x^{2u}-1$ is an odd or even power of x. Hence

$$p_{22}^{\ 1} = y, \quad p_{22}^{\ 2} = \tfrac{1}{2}(v-3) - y,$$

and condition 3 of Theorem 2.1 is satisfied.

Using the identities 1.4 and 1.5, we obtain

$$p_{11}^{\ 1} + p_{12}^{\ 1} = n_1 - 1 = \tfrac{1}{2}(v-3), \quad p_{12}^{\ 1} = y,$$

and

$$p_{21}^{\ 1} + p_{22}^{\ 1} = n_1 = \tfrac{1}{2}(v-1), \quad p_{21}^{\ 1} = \tfrac{1}{2}(v-1) - y.$$

Thus $y = \frac{1}{4}(v-1) = p_{12}^{\ 1} = p_{21}^{\ 1}$; therefore P_1 and P_2 are as stated in the theorem.

If $a = 0$, the resulting series is the completely balanced series given in (6, Theorem 5.1).

Example. If $m = 1, \lambda = 3$, we take $x = 2$ and obtain a design for 53 varieties in blocks of 13 by adding the integers modulo 53 to the initial block

(1, 16, 44, 15, 28, 24, 13, 49, 42, 36, 46, 47, 10).

THEOREM 3.2. *If* $v = 4m(4\lambda - 1) + 1 = p^e$ *and if among the* $2\lambda - 1$ *expressions*

$$x^{4ms} - 1 = x^{q_s} \qquad (s = 1, 2, \ldots, 2\lambda - 1)$$

there are $\lambda - a$ *even and* $\lambda + a - 1$ *odd powers of* x, *then the design with parameters*

$$v = 4m(4\lambda - 1) + 1,\ b = mv,\ r = 4m\lambda,\ k = 4\lambda;$$
$$n_1 = n_2 = 2m(4\lambda - 1),\ \lambda_1 = \lambda - a + 1,\ \lambda_2 = \lambda + a - 1;$$
$$P_1, P_2 \text{ (cf. 3.1)}$$

can be constructed from the initial blocks

$$(0, x^{2i}, x^{2i+4m}, \ldots, x^{2i+4m(4\lambda-2)}),$$

where i *ranges from* 0 *to* $m - 1$.

Proof. The differences involving the zero element are all distinct and cover the even powers of x once. Hence, from Theorem 3.1, $\lambda_1 = \lambda - a + 1$, $\lambda_2 = \lambda + a - 1$; also $n_1 = n_2 = \frac{1}{2}(v - 1) = 2m(4\lambda - 1)$. P_1 and P_2 are obtained as before, and have the same form.

If $a = 1$, a completely balanced series is obtained.

4. Series with $2am(2a\lambda \pm 1) + 1 = p^e$.

THEOREM 4.1. *If* $v = 2am(2a\lambda + 1) + 1 = p^e$, *and if among the exponents* q_s, *where*

$$x^{q_s} = x^{2ams} - 1 \qquad (s = 1, 2, \ldots, a\lambda),$$

the residue class of $(i - 1)$ *modulo* a *is represented* λ_i *times, then the design with parameters*

$$v = 2am(2a\lambda + 1) + 1,\ b = mv,\ r = m(2a\lambda + 1),\ k = 2a\lambda + 1;$$
$$n_i = 2m(2a\lambda + 1),\ \lambda_i;\ P_i,$$

can be constructed from the initial blocks

$$(x^{au}, x^{au+2am}, x^{au+4am}, \ldots, x^{au+4ma^2\lambda}),$$

where u *ranges from* 0 *to* $m - 1$. *The* $p_{ij}^{\,k}$ *are the number of expressions*

$$x^{a^2t+i-j} - 1 = x^z,$$

where $z \equiv k - j \pmod a$, *and* t *ranges from* 0 *to* $(v - a - 1)/a$.

Proof. The differences are

$$x^{au+2arm+q_s},\ x^{au+am(2r+2a\lambda+1-2s)+q_s}.$$

Let set i be the set of powers which are congruent to $(i - 1)$ modulo a. There are a such sets; hence $n_i = (v - 1)/a = 2m(2a\lambda + 1)$. Also, since each element of set i occurs among the differences λ_i times, condition 2 of Theorem 2.1 is satisfied.

The number of times that every element of set k occurs among the differences in (set i − set j) is p_{ij}^{k}; but (set i − set j) consists of elements of the form

$$x^{ai+a\omega+i-1} - x^{a\omega+j-1} = x^{a\omega+j-1}(x^{ai+i-j} - 1) = x^{a\omega+j-1+z},$$

where

$$x^{ai+i-j} - 1 = x^{z}.$$

These elements are in set k if and only if

$$j - 1 + z \equiv k - 1 \pmod{a},$$

that is,

$$z \equiv k - j \pmod{a}.$$

The work of finding the p_{ij}^{k} can be simplified by noting that

$$x^{z_1} = x^{2am(2a\lambda+1)-a} - 1 = (x^{a} - 1)\, x^{am(2a\lambda+1)-a} = x^{z+am(2a\lambda+1)-a},$$

that is, $z \equiv z_1 \pmod{a}$. Hence we need the expressions

$$x^{z} = x^{ai} - 1$$

only for $i = 0, 1, \ldots, (v-1)/2a$. The residue class of z corresponding to a given i is represented twice if $i < (v-1)/2a$ and is represented once if $i = (v-1)/2a$.

Example. Take $a = 3$, $m = \lambda = 1$; then $v = b = 43$, $r = k = 7$, $n_1 = n_2 = n_3 = 14$. From the equations in GF(43)

$$3^{3} - 1 = 3^{17},\quad 3^{6} - 1 = 3^{22},\quad 3^{9} - 1 = 3^{34},\quad 3^{12} - 1 = 3,\quad 3^{15} - 1 = 3^{36},$$
$$3^{18} - 1 = 3^{23},\ 3^{21} - 1 = 3^{4},$$

we deduce that the q_s are 22, 1, and 23, that is, 1, 1, and 2 (mod 3). Hence $\lambda_1 = 0$, $\lambda_2 = 2$, $\lambda_3 = 1$.

To find the p_{ij}^{k}, take $i \equiv j \pmod 3$; p_{ii}^{k} is the number of expressions $x^{3i} - 1 = x^{z}$, where $z \equiv k - i \pmod 3$, and $i = 0, 1, \ldots, 7$. From the preceding tabulation of these expressions, it is seen that $z \equiv 0$ once for $i < 7$ and once for $i = 7$; $z \equiv 1$ three times; $z \equiv 2$ twice. Hence

$$p_{11}^{1} = p_{22}^{2} = p_{33}^{3} = 3;\ p_{11}^{2} = p_{22}^{3} = p_{33}^{1} = 6;\ p_{11}^{3} = p_{22}^{1} = p_{33}^{2} = 4.$$

The remaining p's can be obtained from the relations 1.4, 1.5, and 1.6; thus

$$P_1 = \begin{pmatrix} 3 & 6 & 4 \\ 6 & 4 & 4 \\ 4 & 4 & 6 \end{pmatrix},\quad P_2 = \begin{pmatrix} 6 & 4 & 4 \\ 4 & 3 & 6 \\ 4 & 6 & 4 \end{pmatrix},\quad P_3 = \begin{pmatrix} 4 & 4 & 6 \\ 4 & 6 & 4 \\ 6 & 4 & 3 \end{pmatrix}.$$

In case $a = 1$, there is only one residue class modulo a; hence all the q_s fall in this single residue class and there is only one $\lambda_i = \lambda$. The resulting series is the completely balanced Series B of **(6)**.

THEOREM 4.2. *If $v = 2am(2a\lambda - 1) + 1 = p^{e}$, and if among the exponents q_s, where*

$$x^{q_s} = x^{2ams} - 1, \qquad (s = 1, 2, \ldots, a\lambda - 1),$$

the residue class of $(i-1)$ *modulo* a *is represented* λ_i *times, then the design with parameters*

$$v = 2am(2a\lambda - 1) + 1,\ b = mv,\ r = 2am\lambda,\ k = 2a\lambda;$$
$$n_i = 2m(2a\lambda - 1),\ \lambda_i + \delta_{i1};\quad P_i$$

can be constructed from the initial blocks

$$(0, x^{au}, x^{au+2am}, \ldots, x^{au+2am(2a\lambda-2)}),$$

where u *ranges from* 0 *to* $m-1$.

Proof. Since the class of zero is represented once by all differences involving the zero element, the frequency of occurrence of the residue class of zero among the q_s is $\lambda_1 + 1$. The $p_{ij}{}^k$ are found just as in Theorem 4.1.

5. Series with $am(a\lambda \pm 1) + 1 = p^e$.

THEOREM 5.1. *If* $v = am(a\lambda + 1) + 1 = p^e$, *where* a *is odd, and if among the exponents* q_s, *where*

$$x^{am s} - 1 = x^{q_s} \qquad (s = 1, 2, \ldots, a\lambda)$$

the residue class of $(i-1)$ *modulo* a *is represented* λ_i *times, then the design with parameters:*

$v = am(a\lambda+1)+1,\ b = mv,\ r = m(a\lambda+1),\ k = a\lambda+1;\ n_i = m(a\lambda+1),\ \lambda_i;\ P_i$

can be constructed from the initial blocks

$$(x^{au}, x^{au+am}, x^{au+2am}, \ldots, x^{au+ma^2\lambda}),$$

where u *ranges from* 0 *to* $m-1$.

Proof. The proof is similar to that of Theorem 4.1; the P_i can be found as before, except that here, since a is odd, there are no relations among the q_s to simplify the work.

THEOREM 5.2. *If* $v = am(a\lambda - 1) + 1 = p^e$, *where* a *is odd, and if among the exponents* q_s, *where*

$$x^{am s} - 1 = x^{q_s} \qquad (s = 1, 2, \ldots, a\lambda - 2)$$

the residue class of $(i-1)$ *modulo* a *is represented* λ_i *times, then the design with parameters*

$v = am(a\lambda - 1) + 1,\ b = mv,\ r = am\lambda,\ k = a\lambda;\ n_i = m(a\lambda - 1),\ \lambda_i + 2\delta_{i1};\ P_i$

can be constructed from the initial blocks

$$(0, x^{au}, x^{au+am}, \ldots, x^{au+am(a\lambda-2)})$$

where u *ranges from* 0 *to* $m-1$.

Proof. The proof is similar to that of Theorem 4.2.

6. Series with $ma + 1 = p^e$. To every element y^f of GF(p^e), let there correspond n varieties

$$y_1{}^f, y_2{}^f, \ldots, y_n{}^f.$$

Then designs with three associate classes can be formed by taking as first associates of any variety

$$y_u{}^{f_1}$$

all varieties

$$y_u{}^{f_2} \qquad (f_1 \neq f_2);$$

the second associates are all varieties

$$y_w{}^{f_1} \qquad (u \neq w);$$

the third associates are all varieties

$$y_w{}^{f_2}.$$

Thus the first associates of a variety are all varieties giving rise to pure differences; the second associates are all varieties giving rise to zero mixed differences; the third associates are all other varieties giving rise to non-zero mixed differences.

THEOREM 6.1. *If $ma + 1 = p^e$, then the design with parameters*

$v = n(ma + 1)$, $b = \frac{1}{2}mv(n - 1)$, $r = am(n - 1)$, $k = 2a$;
$n_1 = ma$, $n_2 = n - 1$, $n_3 = ma(n - 1)$, $\lambda_1 = (a - 1)(n - 1)$, $\lambda_2 = ma$, $\lambda_3 = a - 1$;

$$P_1 = \begin{pmatrix} ma - 1 & 0 & 0 \\ 0 & 0 & n - 1 \\ 0 & n - 1 & (ma - 1)(n - 1) \end{pmatrix},$$

$$P_2 = \begin{pmatrix} 0 & 0 & ma \\ 0 & n - 2 & 0 \\ ma & 0 & ma(n - 2) \end{pmatrix},$$

$$P_3 = \begin{pmatrix} 0 & 1 & ma - 1 \\ 1 & 0 & n - 2 \\ ma - 1 & n - 2 & (ma - 1)(n - 2) \end{pmatrix}$$

can be constructed from the initial blocks

$$(x_u{}^i, x_u{}^{i+m}, \ldots, x_u{}^{i+(a-1)m}, x_w{}^i, x_w{}^{i+m}, \ldots, x_w{}^{i+(a-1)m}),$$

where i ranges from 0 to $m - 1$, and (u, w) are the $\frac{1}{2}n(n - 1)$ pairs of integers selected from 1 to n.

Proof. The pure differences of the type (u, u), arising from all blocks with a fixed u and w, are each repeated $a - 1$ times (**6**, Theorem 2.1). Since a fixed u occurs with $n - 1$ values of w, $\lambda_1 = (n - 1)(a - 1)$. The zero mixed differences of the type (u, w) occur a times for i fixed, and ma times in all; hence

$\lambda_2 = ma$. The non-zero mixed differences of the type (u, w) occur $\lambda_3 = a - 1$ times (6, Theorem 2.1). Since there are $\frac{1}{2}n(n-1)$ pairs (u, w),

$$b = \tfrac{1}{2}m(ma+1)\, n(n-1).$$

The number of first associates of a variety $y_u^{f_1}$ is the number of varieties of the form

$$y_u^{f_2};$$

thus $n_1 = ma$. Also, n_2 is the number of varieties of the form

$$y_w^{f_1},$$

that is, $n_2 = n - 1$; $n_3 = ma(n-1)$ = the number of varieties of the form

$$y_w^{f_2}.$$

Two first associates have the form

$$y_u^{f_1}, y_u^{f_2};$$

hence p_{11}^1 is the number of varieties of the form

$$y_u^{f_2} \qquad (f_1 \neq f_2, f_2 \neq f_3),$$

that is, $ma - 1$. Also, $p_{12}^1 = p_{13}^1 = p_{22}^1 = 0$. We obtain p_{23}^1 as the number of varieties of the form

$$y_w^{f_1} \qquad (f_1 \text{ fixed});$$

hence $p_{23}^1 = n - 1$. The number of varieties of the form

$$y_w^{f_2} \qquad (u \neq w)$$

gives us $p_{33}^1 = (ma-1)(n-1)$. This completes P_1 and the matrices P_2 and P_3 can be found in a similar way.

THEOREM 6.2. *If $ma + 1 = p^e$, then the design with parameters*

$v = n(ma+1)$, $b = m(n-1)v$, $r = m(a+1)(n-1)$, $k = a+1$;
$n_1 = am$, $n_2 = n-1$, $n_3 = ma(n-1)$, $\lambda_1 = (a-1)(n-1)$, $\lambda_2 = 0$, $\lambda_3 = 2$;
P_1, P_2, P_3 (cf. 6.1)

can be constructed from the initial blocks

$$(x_u^i, x_u^{i+m}, \ldots, x_u^{i+(a-1)m}, 0_w),$$

where i ranges from 0 to $m-1$ and (u, w) are the $n(n-1)$ permutations of the integers from 1 to n, taken two at a time.

Proof. Each pure difference of the type (u, u) occurs $a - 1$ times for w fixed (6, Theorem 2.1) and $(n-1)(a-1)$ times in all. The non-zero mixed differences of the type (u, w) occur twice, since u and w can be interchanged; the zero mixed differences do not occur at all. The n_i and P_i are found as in Theorem 6.1 and have the same values.

COROLLARY 6.21. *If we permit $u = w$, we obtain a series with parameters*

$v = n(am + 1)$, $b = mnv$, $r = mn(a + 1)$, $k = a + 1$; n_i (cf. 6.2), $\lambda_1 = n(a - 1) + 2$, $\lambda_2 = 0$, $\lambda_3 = 2$; P_i (cf. 6.2).

Proof. Each pure difference of the type (u, u) arises $a + 1$ times from the blocks in which $u = w$.

COROLLARY 6.22. *If $n = a + 1$ in Theorem* 6.2 *and the block $(0_1, 0_2, \ldots, 0_{a+1})$ is included twice in the set of initial blocks, we obtain a group-divisible design with parameters*

$v = (a + 1)(am + 1)$, $b = \{m(a + 1) + 2\}(am + 1)$, $r = ma(a + 1) + 2$, $k = a + 1$;
$g = a + 1$, $h = am + 1$; $\lambda_1 = a(a - 1)$, $\lambda_2 = 2$,

where g is the number of groups in the GD *design, h is the number of treatments in a group.*

Proof. To combine the second and third classes of a PBIBD, the following conditions must hold:

(1) $\lambda_2 = \lambda_3$,
(2) $p_{13}^2 + p_{12}^2 = p_{13}^3 + p_{12}^3$,
(3) $p_{22}^2 + p_{33}^2 + p_{23}^2 + p_{32}^2 = p_{22}^3 + p_{33}^3 + p_{23}^3 + p_{32}^3$.

These conditions are satisfied; so classes 2 and 3 can be combined to give a PBIBD with the stated parameters. For this design,

$$n_1 = ma,\ n_2 = (ma + 1)(n - 1);$$

$$P_1 = \begin{pmatrix} ma - 1 & 0 \\ 0 & (ma + 1)a \end{pmatrix}, \qquad P_2 = \begin{pmatrix} 0 & ma \\ ma & (ma + 1)(a - 1) \end{pmatrix}.$$

COROLLARY 6.23. *If $n = a + 1$ in Corollary* 6.21 *and the block $(0_1, 0_2, \ldots, 0_{a+1})$ is included twice in the set of initial blocks, we obtain a* GD *design with parameters*

$v = (am + 1)(a + 1)$, $b = \{m(a + 1)^2 + 2\}(ma + 1)$, $r = (a + 1)^2m + 2$, $k = a + 1$;
$g = a + 1$, $h = am + 1$; $\lambda_1 = a^2 + 1$, $\lambda_2 = 2$.

Proof. Similar to that of Corollary 6.22.

COROLLARY 6.24. *If in Theorem* 6.2 *the initial blocks are replaced by*

$$(0_u, x_u^{\,i}, x_u^{\,i+m}, \ldots, x_u^{\,i+(a-1)m}, 0_w) \quad (i = 0, 1, \ldots, m - 1)$$

where (u, w) are permutations of the integers from 1 to n, taken two at a time, then the resulting design is a GD *design with parameters*

$v = n(am + 1)$, $b = m(n - 1)v$, $r = (a + 2)(n - 1)m$, $k = a + 2$;
$g = n$, $h = am + 1$; $\lambda_1 = (a + 1)(n - 1)$, $\lambda_2 = 2$.

Proof. Here each zero mixed difference of the type (u, w) also occurs twice; hence classes 2 and 3 can be combined as in Corollary 6.22.

Example. Take $a = 2, m = 1$, in Corollary 6.23; the GD design has parameters $v = 9, b = 33, r = 11, k = 3; g = h = 3; \lambda_1 = 5, \lambda_2 = 2$. The initial blocks are $(1, 2, 0)$, $(1, 2, 3)$, $(1, 2, 6)$, $(4, 5, 0)$, $(4, 5, 3)$, $(4, 5, 6)$, $(7, 8, 0)$, $(7, 8, 3)$, $(7, 8, 6)$, $(0, 3, 6)$, $(0, 3, 6)$, where we set $y_1 = y$, $y_2 = y + 3$, $y_3 = y + 6$. The three groups of three varieties are 0, 1, 2; 3, 4, 5; and 6, 7, 8. Thus, for instance, 3 occurs five times with 4 and 5 and twice with all the other varieties.

7. Series with $2m(2\lambda \pm 1) + 1 = p^e$.

THEOREM 7.1. *If $2m(2\lambda + 1) + 1 = p^e$, then the design with parameters*

$v = \{2m(2\lambda + 1) + 1\}n, b = \frac{1}{2}(n - 1)mv, r = m(2\lambda + 1)(n - 1)$,
$k = 4\lambda + 2$;
$n_1 = 2m(2\lambda + 1), n_2 = n - 1, n_3 = n_1(n - 1), \lambda_1 = (n - 1)\lambda$,
$\lambda_2 = m(2\lambda + 1), \lambda_3 = \lambda$;

$$P_1 = \begin{pmatrix} n_1 - 1 & 0 & 0 \\ 0 & 0 & n - 1 \\ 0 & n - 1 & (n_1 - 1)(n - 1) \end{pmatrix},$$

$$P_2 = \begin{pmatrix} 0 & 0 & n_1 \\ 0 & n - 2 & 0 \\ n_1 & 0 & n_1(n - 2) \end{pmatrix},$$

$$P_3 = \begin{pmatrix} 0 & 1 & n_1 - 1 \\ 1 & 0 & n - 2 \\ n_1 - 1 & n - 2 & (n_1 - 1)(n - 2) \end{pmatrix}$$

can be constructed from the initial blocks

$$(x_u^{\,i}, x_u^{\,i+2m}, \ldots, x_u^{\,i+4m\lambda}, x_w^{\,i}, x_w^{\,i+2m}, \ldots, x_w^{\,i+4m\lambda}),$$

where i ranges from 0 to $m - 1$ and (u, w) are the $\frac{1}{2}n(n - 1)$ pairs of integers selected from 1 to n.

Proof. By (6, Theorem 3.1), the pure differences of the type (u, u) occur λ times for a fixed w and $\lambda_1 = (n - 1)\lambda$ times in all. Similarly, the non-zero mixed differences of the type (u, w) occur $\lambda_3 = \lambda$ times while the zero mixed differences of the type (u, w) occur $(2\lambda + 1)$ times for a fixed i, that is, $\lambda_2 = m(2\lambda + 1)$ times in all. The method employed in Theorem 6.1 gives the values for the n_i and shows that the P_i have the same form as in that theorem, with ma replaced by $2m(2\lambda + 1)$.

Example. $m = 2, n = 3, \lambda = 1$, gives a design with $v = 39, b = 78$,
$r = 12, k = 6$; $n_1 = 12, n_2 = 2, n_3 = 24, \lambda_1 = 2, \lambda_2 = 6, \lambda_3 = 1$;

$$P_1 = \begin{pmatrix} 11 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 2 & 22 \end{pmatrix}, \quad P_2 = \begin{pmatrix} 0 & 0 & 12 \\ 0 & 1 & 0 \\ 12 & 0 & 12 \end{pmatrix}, \quad P_3 = \begin{pmatrix} 0 & 1 & 11 \\ 1 & 0 & 1 \\ 11 & 1 & 11 \end{pmatrix}.$$

The initial blocks are

$$(2_u^i, 2_u^{i+4}, 2_u^{i+8}, 2_w^i, 2_w^{i+4}, 2_w^{i+8}),$$

that is, reduced modulo 13,

$$(i+1)(1_u, 3_u, 9_u, 1_w, 3_w, 9_w), \qquad i = 0 \text{ or } 1.$$

The pairs (u, w) are the pairs $(1, 2)$, $(1, 3)$, and $(2, 3)$. Setting $y_1 = y$; for $y = 0, 1, \ldots, 12$; $y_2 = y + 13$; $y_3 = y + 26$; we find that the initial blocks are

$$(1, 3, 9, 14, 16, 22), \quad (1, 3, 9, 27, 29, 35), \quad (2, 6, 5, 15, 19, 18),$$
$$(2, 6, 5, 28, 32, 31), \quad (14, 16, 22, 27, 29, 35), \quad (15, 19, 18, 28, 32, 31).$$

The other blocks are generated by addition modulo 13; thus the first block generates $(2, 4, 10, 15, 17, 23)$, $(3, 5, 11, 16, 18, 24), \ldots,$ $(0, 2, 8, 13, 15, 21)$. Given any variety, say 1, its first associates are $0, 2, 3, \ldots, 12$; its second associates are 14 and 27; the remaining varieties are third associates.

THEOREM 7.2. *If* $2m(2\lambda + 1) + 1 = p^e$, *then the design with parameters*

$$v = \{2m(2\lambda + 1) + 1\}n, b = (n - 1)mv, r = m(2\lambda + 2)(n - 1),$$
$$k = 2\lambda + 2; n_1 \text{ (cf. 7.1)}, \lambda_1 = (n - 1)\lambda, \lambda_2 = 0, \lambda_3 = 1; P_1 \text{ (cf. 7.1)}$$

can be constructed from the initial blocks

$$(x_u^i, x_u^{i+2m}, \ldots, x_u^{i+4m\lambda}, 0_w),$$

where i *ranges from* 0 *to* $m - 1$ *and* (u, w) *are the* $n(n - 1)$ *permutations of the integers from* 1 *to* n, *taken two at a time.*

Proof. Proceed as in Theorem 6.2.

Using proofs similar to those used in the Corollaries to Theorem 6.2, we obtain

COROLLARY 7.21. *If we permit* $u = w$ *in Theorem* 7.2, *we obtain a series with*

$$v = \{2m(2\lambda + 1) + 1\}n, \ b = mnv, \ r = mn(2\lambda + 2), \ k = 2\lambda + 2;$$
$$n_1 = 2m(2\lambda+1), n_2 = n-1, n_3 = n_1(n-1), \ \lambda_1 = n\lambda+1, \lambda_2 = 0, \lambda_3 = 1.$$

COROLLARY 7.22. *If in Theorem* 7.2 *we set* $n = 2\lambda + 2$ *and include the block*

$$(0_1, 0_2, \ldots, 0_{2\lambda+2})$$

among the initial blocks, then we obtain a GD *design with parameters*

$$v = (2\lambda + 2)\{2m(2\lambda + 1) + 1\},$$
$$b = \{m(2\lambda + 2)(2\lambda + 1) + 1\}\{2m(2\lambda + 1) + 1\},$$
$$r = (2\lambda + 2)(2\lambda + 1)m + 1, \ k = 2\lambda + 2;$$
$$g = 2\lambda + 2, \ h = 2m(2\lambda + 1) + 1; \ \lambda_1 = \lambda(2\lambda + 1), \lambda_2 = 1.$$

COROLLARY 7.23. *If in Corollary* 7.21 *we set* $n = 2\lambda + 2$ *and include the initial block*

$$(0_1, 0_2, \ldots, 0_{2\lambda+2}),$$

then we obtain a GD *design with parameters*

$v = (2\lambda + 2)\{2m(2\lambda + 1) + 1\}$,
$b = \{m(2\lambda + 2)^2 + 1\}\{2m(2\lambda + 1) + 1\}$,
$r = (2\lambda + 2)^2m + 1, k = 2\lambda + 2$;
$g = 2\lambda + 2, h = 2m(2\lambda + 1) + 1$; $\lambda_1 = (2\lambda + 2)\lambda + 1, \lambda_2 = 1$.

THEOREM 7.3. *If* $2m(2\lambda - 1) + 1 = p^e$, *then the design with parameters*

$v = \{2m(2\lambda - 1) + 1\}n, b = \frac{1}{2}mv(n - 1), r = 2m(n - 1)\lambda, k = 4\lambda$;
$n_1 = 2m(2\lambda - 1), n_2 = n - 1, n_3 = n_1(n - 1)$,
$\lambda_1 = (n - 1)\lambda, \lambda_2 = 2m\lambda, \lambda_3 = \lambda$;

$$P_1 = \begin{pmatrix} n_1 - 1 & 0 & 0 \\ 0 & 0 & n - 1 \\ 0 & n - 1 & (n_1 - 1)(n - 1) \end{pmatrix},$$

$$P_2 = \begin{pmatrix} 0 & 0 & n_1 \\ 0 & n - 2 & 0 \\ n_1 & 0 & n_1(n - 2) \end{pmatrix},$$

$$P_3 = \begin{pmatrix} 0 & 1 & n_1 - 1 \\ 1 & 0 & n - 2 \\ n_1 - 1 & n - 2 & (n_1 - 1)(n - 2) \end{pmatrix}$$

can be constructed from the initial blocks

$$(0_u, x_u^i, x_u^{i+2m}, \ldots, x_u^{i+4(\lambda-1)m}, 0_w, x_w^i, x_w^{i+2m}, \ldots, x_w^{i+4(\lambda-1)m}),$$

where i *ranges from* 0 *to* $m - 1$ *and* (u, w) *are the* $\frac{1}{2}n(n - 1)$ *pairs of integers selected from* 1 *to* n.

Proof. Proceed as in Theorem 7.1.

THEOREM 7.4. *If* $2m(2\lambda - 1) + 1 = p^e$, *then the design with parameters*

$v = \{2m(2\lambda-1)+1\}n, b = mv(n-1), r = (2\lambda+1)m(n-1), k = 2\lambda+1$;
n_i (cf. 7.3), $\lambda_1 = (n - 1)\lambda, \lambda_2 = 2, \lambda_3 = 1$; P_i (cf. 7.3)

can be constructed from the initial blocks

$$(0_u, x_u^i, x_u^{i+2m}, \ldots, x_u^{i+4m\lambda}, 0_w),$$

where i *ranges from* 0 *to* $m - 1$ *and* (u, w) *are the* $n(n - 1)$ *permutations of the integers from* 1 *to* n, *taken two at a time.*

Proof. Proceed as in Theorem 7.2.

References

1. R. C. Bose, W. H. Clatworthy, and S. S. Shrikhande, *Tables of partially balanced designs with two associate classes* (University of North Carolina, 1954).
2. R. C. Bose and W. S. Connor, *Combinatorial properties of group divisible incomplete block designs*, Ann. Math. Statist., *23* (1952), 367–387.
3. R. C. Bose and K. R. Nair, *Partially balanced incomplete block designs*, Sankhya, *4* (1938), 337–372.
4. R. C. Bose and T. Shimamoto, *Classification and analysis of partially balanced designs with two associate classes*, J. Amer. Statist. Assoc., *47* (1952), 151–184.
5. R. C. Bose, S. S. Shrikhande, and K. Bhattacharya, *On the construction of group divisible incomplete block designs*, Ann. Math. Statist., *24* (1953), 161–195.
6. D. A. Sprott, *A note on balanced incomplete block designs*, Can. J. Math., *6* (1954), 341–346.

University of Toronto

A CLASS OF ALGEBRAS WITHOUT UNITY ELEMENT

R. M. THRALL

1. Introduction. In a study of the commuting algebra of tensor space representations of the orthogonal group W. P. Brown encountered a class of algebras for which the existence of a unity element was equivalent to semisimplicity, but which were of interest whether or not semisimple. He gave these algebras the name *generalized-total matrix algebras* and proved **(2)** that each such algebra was characterized by three integers l, r, m and was isomorphic to the algebra of all square matrices of degree $r + l + m$ which have zeros in the first l rows and in the last r columns.

Let F be any field, let K be an extension sfield of finite degree k over F, let m be a positive integer and let l, r be non-negative integers. We denote by $C = C(K, m, l, r)$ the F-algebra of order $k(m + l)(m + r)$ consisting of all K-matrices having zeros in the first l rows and in the last l-columns. We call C a *submatrix algebra.*

In the present paper we introduce a new family of algebras called *algebras of class Q.* These algebras are defined in terms of certain simple properties possessed by submatrix algebras. Our main result is a proof that each algebra of class Q is a factor algebra of a direct sum of submatrix algebras. We also touch on the topics of automorphisms, isomorphisms, and representations, of algebras of class Q.

2. Algebras of class Q and class Q'. An F-algebra A (of finite dimension) is said to be of class Q if there exists an idempotent ϵ in A such that the following three conditions hold:

(Q_1) $B = \epsilon A \epsilon$ is semisimple,
(Q_2) $A \epsilon A = A$,
(Q_3) $A = B + N$, where N is the radical of A.

If instead of (Q_1) we have the stronger condition

(Q_1') $B = \epsilon A \epsilon$ is simple,

then we say that A is of *class Q'*.

It is easy to see that every submatrix algebra $C(K, m, l, r)$ is of class Q'; for we may take as the idempotent the matrix

$$\epsilon' = \begin{Vmatrix} 0 & 0 & 0 \\ 0 & I_m & 0 \\ 0 & 0 & 0 \end{Vmatrix} \tag{1}$$

where the partitioning of rows and columns is given by l, m, r.

Received October 12, 1954. Part of this research was carried out under a contract of the Office of Ordnance Research with the University of Michigan.

THEOREM 1. *Let A be an algebra of class Q. Then $ANA = 0$.*

Since no non-zero element in a semisimple algebra can be in the radical we note that $B \cap N = \{0\}$; hence the sum $B + N$ is direct (in the vector space sense). Now multiply A on left and right by ϵ and we get from (Q_3) that

$$\epsilon N \epsilon = 0. \tag{2}$$

Finally, we have

$$ANA = A\epsilon ANA\epsilon A \subseteq A\epsilon N\epsilon A = 0.$$

COROLLARY 1. *If A is of class Q then $N^3 = 0$ and $AN^2 = N^2A = 0$.*

THEOREM 2. *If A is of class Q then*

$$A = B + \epsilon N + N\epsilon + N^2 \quad \text{(direct sum)}. \tag{3}$$

We establish the theorem by identifying (3) with the Pierce decomposition of A relative to ϵ. By (Q_1), $B = \epsilon A \epsilon$ consists of all elements of A having ϵ as two-sided unity. Next, suppose that α is an element of A for which $\alpha = \epsilon\alpha$ and $\alpha\epsilon = 0$. According to (Q_3) we can write $\alpha = \beta + \eta$ where $\beta \in B$ and $\eta \in N$. Now since $\alpha = \epsilon\alpha$ we must have $\epsilon\eta = \eta$, and then $\alpha\epsilon = \beta + \epsilon\eta\epsilon = \beta$ requires $\beta = 0$. This shows that ϵN contains all elements of A having ϵ as left unity and right annihilator; moreover, it follows from (2) that each element of ϵN has this property. We show similarly that $N\epsilon$ consists of all elements of A having ϵ as right unity and left annihilator.

This shows that the Pierce decomposition is

$$A = B + \epsilon N + N\epsilon + N_0 \quad \text{(direct sum)} \tag{4}$$

where N_0 consists of all elements of A having ϵ as two-sided annihilator. All that remains is to show that $N_0 = N^2$. It is a consequence of Corollary 1 that $N^2 \subseteq N_0$. Next, it follows from (Q_2), (4), and the fact that N is an ideal that

$$\begin{aligned} A = A \cdot A &= (B + N\epsilon)(B + \epsilon N) \\ &= B + N\epsilon + \epsilon N + N\epsilon N, \end{aligned} \tag{5}$$

where $N\epsilon N \subseteq N^2 \subseteq N_0$. Comparison of (4) and (5) then shows that $N\epsilon N = N_0$ and hence

$$N^2 = N\epsilon N = N_0.$$

3. Structure theory. In this section we show how a general algebra of class Q can be built up from algebras of class Q' and in the following section we study the structure of algebras of class Q'.

THEOREM 3. *Let A, A_1, and A_2 be algebras of class Q and let M be an ideal (two-sided) in A. Then the direct sum $A_1 \oplus A_2$ and the residue class algebra $A - M$ are both of class Q.*

If ϵ, ϵ_1, ϵ_2 are the postulated idempotents in A, A_1, A_2, respectively, then (ϵ_1, ϵ_2) and $\epsilon + M$, respectively, are idempotents which satisfy (Q_1) and (Q_2) in $A_1 + A_2$ and $A - M$. It is also easy to check (Q_3) in both cases.

THEOREM 4. *Let A be an algebra of class Q. Then $A = A^* - M^*$ where A^* is a direct sum of algebras of class Q' and M^* is contained in the square of the radical of A^*.*

It follows from Theorem 3 that $A^* - M^*$ is of class Q if A^* and M^* satisfy the conditions of Theorem 4.

Let $A = B + N$ where B and N are related to A as in (Q_1), (Q_2), (Q_3), and suppose that

$$B = B_1 + \ldots + B_p \tag{6}$$

is the (unique) expression of B as a direct sum of simple subalgebras. Let ϵ_i be the unity element of B_i $(i = 1, \ldots, p)$; then

$$\epsilon = \epsilon_1 + \ldots + \epsilon_p \tag{7}$$

is an expression of ϵ as a sum of orthogonal idempotents in the centre of B.

Now set

$$A_i = A\epsilon_i A \qquad (i = 1, \ldots, p). \tag{8}$$

and

$$N_i = A_i \cap N \qquad (i = 1, \ldots, p). \tag{9}$$

LEMMA 1. *A_i is of class Q' with idempotent ϵ_i, simple summand B_i, and radical N_i.*

The equation

$$\epsilon_i A_i \epsilon_i = \epsilon_i(A\epsilon_i A)\epsilon_i = \epsilon_i \epsilon A \epsilon\epsilon_i \epsilon A \epsilon\epsilon_i = \epsilon_i B \epsilon_i B \epsilon_i = B_i B_i = B_i$$

verifies (Q_1').

Next, we have

$$A_i \supseteq A_i \epsilon_i A_i = A\epsilon_i A \epsilon_i A \epsilon_i A \supseteq A\epsilon_i A = A_i,$$

and hence $A_i = A_i \epsilon_i A_i$. This verifies (Q_2).

Finally,

$$A_i = (B + N)\epsilon_i(B + N) = B\epsilon_i B + (N\epsilon_i B + N\epsilon_i N + B\epsilon_i N) \subseteq B_i + N_i$$

since A_i and N_i are ideals in A. But $B_i \subseteq A_i$, $N_i \subseteq A_i$, hence

$$A_i = B_i + N_i. \tag{10}$$

This sum is direct (in the vector space sense) since ϵ_i is unity element for B_i and $\epsilon_i N_i \epsilon_i \subseteq \epsilon_i N \epsilon_i = \epsilon_i \epsilon N \epsilon\epsilon_i = 0$; it follows that $A_i - N_i \cong B_i$, and hence the N_i is the radical of A_i. This establishes (Q_3).

LEMMA 2. *$A_i A_j = 0$ if $i \neq j$.*

If $i \neq j$ we have

$$\begin{aligned} \epsilon_i A\, \epsilon_j &= \epsilon_i B\, \epsilon_j + \epsilon_i N\, \epsilon_j \\ &= B\, \epsilon_i\, \epsilon_j + \epsilon_i\, \epsilon N \epsilon\, \epsilon_j \\ &= 0 + 0; \end{aligned}$$

hence $A_i A_j = A\epsilon_i A A \epsilon_j A = 0$.

We are now ready to prove Theorem 4. Let A^* be the (ring) direct sum of $A_1, \ldots, A_p$, i.e., A^* consists of all p-tuples $\alpha^* = (\alpha_1, \ldots, \alpha_p)$ with α_i in A_i and with addition and multiplication done componentwise. Consider the mapping

$$T : \alpha^* = (\alpha_1, \ldots, \alpha_p) \to \alpha = \alpha_1 + \ldots + \alpha_p. \tag{11}$$

of A^* into A.

This mapping is clearly a linear transformation. It follows from Lemma 2 that it is a ring homomorphism. It is "onto" since

$$A^*T = A_1 + \ldots + A_p = A\epsilon_1 A + \ldots + A\epsilon_p A = A\epsilon A = A.$$

Let M^* be the kernel of T, and let N^* be the radical of A^*. All that remains is to show that $M^* \subseteq (N^*)^2$. Suppose that $\alpha^* = (\alpha_1, \ldots, \alpha_p)$ lies in M^*;

$$\alpha = \alpha_1 + \ldots + \alpha_h = 0. \tag{12}$$

It follows from (3) that we can write

$$\alpha_i = \beta_i + \eta_i + \zeta_i + \tau_i \qquad (i = 1, \ldots, p) \tag{13}$$

where $\beta_i = \epsilon_i \alpha_i \epsilon_i$, $\eta_i = \epsilon_i \eta_i$, $\zeta_i = \zeta_i \epsilon_i$ and $\eta_i \epsilon_i = \epsilon_i \zeta_i = \epsilon_i \tau_i = \tau_i \epsilon_i = 0$.

Now $\epsilon_i \alpha \epsilon_i = \beta_i$, $\epsilon_i \alpha = \beta_i + \eta_i$, $\alpha\epsilon_i = \beta_i + \zeta_i$, and since $\alpha = 0$ this gives $\beta_i = \eta_i = \zeta_i = 0$ $(i = 1, \ldots, p)$ and hence

$$\alpha^* = (\tau_1, \ldots, \tau_p)$$

which is in $(N^*)^2$.

4. Structure theory (continued). Theorem 4 gives the structure of algebras of class Q in terms of algebras of class Q'. In this section we refine this result by an analysis of algebras of class Q'.

THEOREM 5. *Every algebra of class Q' is the homomorphic image of a submatrix algebra. The kernel of the homomorphism is contained in the square of the radical of the submatrix algebra.*

Let A be an algebra of class Q' with simple summand $B = \epsilon A \epsilon$. According to Wedderburn's Theorem **(1)**, B is isomorphic to a total matrix algebra over a finite extension sfield K of F. Let $\{\kappa_1, \ldots, \kappa_k\}$ be a basis for K over F, let m be the degree of B over K, and let e_{ij} $(i, j = 1, \ldots, m)$ be a matrix unit K-basis for B.

Let $e = e_{11}$; then K is isomorphic to $K' = eBe$ and every irreducible left-B-space W is isomorphic to eB. In particular, $eW \neq 0$ and there exists a vector w in W such that $\{w(= e_{11} w), e_{21} w, \ldots, e_{m1} w\}$ is a K-basis for W. Then the mk vectors $\kappa_h e_{i1} w$ $(h = 1, \ldots, k; i = 1, \ldots, m)$ form an F-basis for W.

ϵN is a left B-space and as such is a direct sum of irreducible left B-spaces. Hence, there exist vectors $\eta_1, \ldots, \eta_l$ in eN such that

$$\{\kappa_h e_{i1} \eta_s \ (h = 1, \ldots, k; i = 1, \ldots, m; s = 1, \ldots, l)\} \tag{14}$$

is an F-basis for ϵN.

Similarly, there exist vectors $\zeta_1, \ldots, \zeta_r$ in Ne such that

$$\{\zeta_t e_{ij} \kappa_h \quad (h = 1, \ldots, k; j = 1, \ldots, m; t = 1, \ldots, r)\} \tag{15}$$

is an F-basis for Ne.

Finally, $N^2 = NeN$ is spanned by all products of basis vectors for Ne and eN; i.e. by all products of the form

$$\zeta_t e_{1j} \kappa_h \kappa_{h'} e_{i1} \eta_s.$$

Next, we observe that the matrix units e_{ij} commute with elements of K, hence $e_{1j} \kappa_h \kappa_{h'} e_{i1}$ is equal to zero unless $i = j$ and then it is equal to $e \kappa_h \kappa_{h'} e$. Now $\zeta_t e = \zeta_t$ and $e\eta_s = \eta_s$; hence N^2 is spanned by the klr products

$$\{\zeta_t \kappa_h \eta_s \quad (h = 1, \ldots, k; s = 1, \ldots, l; t = 1, \ldots, r)\}. \tag{16}$$

In general these products will not be independent but will satisfy certain linear conditions

$$\sum_{t,h,s} a_q(t, h, s)\, \zeta_t \kappa_h \eta_s = 0 \qquad (q = 1, \ldots, g). \tag{17}$$

We may suppose that these conditions are independent; then the F-dimension of N^2 is $klr - g$.

We now let $C = C(K, m, l, r)$. Let ϵ' be the matrix given in (1), and let $C = B' + N'$ be the decomposition (Q_3) for C. Then we can choose matrix units e_{ij}' for B' and elements $\eta_1', \ldots, \eta_l'$ in $e_{11}' N'$, $\zeta_1', \ldots, \zeta_r'$ in $N' e_{11}'$ such that the $k(m + l)(m + r)$ elements

$$\kappa_h e_{ti} \eta_j, \ \kappa_h e_{ti}' \eta_s', \ \zeta_t' e_{ij} \kappa_h, \ \zeta_t' \kappa_h \eta_s' \tag{18}$$

$$(h = 1, \ldots, k; i, j = 1, \ldots, m; s = 1, \ldots, l; t = 1, \ldots, r)$$

form an F-basis for C. Then the unique linear transformation of C onto A which sends each of these basis vectors into the corresponding unprimed element of A is clearly a ring homomorphism whose kernel lies in $(N')^2$. This completes the proof of Theorem 5.

Next we combine the results of Theorems 4 and 5 and get our main structure theorem.

THEOREM 6. *Let A be an algebra of class Q. Then $A = A^* - M^*$ where A^* is a direct sum of submatrix algebras and M^* is contained in the square of the radical of A^*.*

5. Representation theory. Let A be an algebra of class Q and let V be a (left) representation space for A. Consider the chain $V \supseteq AV \supseteq NAV \supseteq ANAV = 0$. Clearly both the spaces V/AV and NAV are annihilated by every element of A, and AV/NAV is a completely reducible non-degenerate A-space. Hence, by suitable choice of basis vectors we get the following matrix form for the representation:

$$\alpha \to V_{(\alpha)} = \begin{Vmatrix} 0 & 0 & 0 \\ V_{21}(\alpha) & V_{22}(\alpha) & 0 \\ V_{31}(\alpha) & V_{32}(\alpha) & 0 \end{Vmatrix} \tag{19}$$

where V_{22} is in completely reduced form. We may suppose the basis elements so chosen that $V_{22}(\epsilon)$ is the identity matrix, and if $\alpha = \beta + \eta + \zeta + \tau$ is a splitting of α according to (3) then

$$V_{22}(\alpha) = V_{22}(\beta),\ V_{21}(\alpha) = V_{21}(\eta),\ V_{32}(\alpha) = V_{32}(\zeta),$$

and

$$V_{31}(\alpha) = V_{31}(\tau).$$

It is easy to show that if any of the integers l_i, r_i defined by the ideals A_i of A given by (8) exceeds unity, then A has unbounded representation type (3). Consequently, it is not likely that there is any simple classification of the indecomposable representations of algebras of class Q.

6. **Uniqueness and automorphisms.** Since the structure theorems depend on the decomposition $A = B + N$ it seems desirable to study its uniqueness. Since N is the radical any lack of uniqueness must come from the semisimple summand B. But since $B = \epsilon A \epsilon$ is uniquely determined by ϵ any second decomposition must correspond to a second idempotent ϵ'.

It is easy to verify for any η_0, ζ_0 in ϵN, $N\epsilon$, respectively, that

$$\epsilon' = \epsilon + \eta_0 + \zeta_0 + \zeta_0\eta_0 = (\epsilon + \zeta_0)(\epsilon + \eta_0) \tag{20}$$

is an idempotent for which Q_1, Q_2, and Q_3 hold. Moreover, if either η_0 or ζ_0 is different from zero, $B' = \epsilon' A \epsilon'$ is not the same as B. Hence all we can expect for B is uniqueness to within an automorphism and this is established in the following theorem.

THEOREM 7. *Let A be an algebra of class Q and let ϵ, ϵ' be idempotents for which Q_1, Q_2, Q_3 hold. Then there is an automorphism T of A which sends ϵ into ϵ'. More precisely, if $\alpha = \beta + \eta + \zeta + \tau$ is a splitting of α according to* (3) *then the mapping*

$$T : \alpha \to \alpha T = \alpha' = \epsilon'\beta\epsilon' + \epsilon'\eta + \zeta\epsilon' + \tau$$

is an automorphism of A which sends ϵ into ϵ'.

Both ϵ and ϵ' are mapped into the identity element of $A - N$ under the natural mapping; hence $\epsilon' - \epsilon$ is in N and so from (3) we get

$$\epsilon' = \epsilon + \eta_0 + \zeta_0 + \tau_0.$$

Now $\epsilon' = (\epsilon')^2 = \epsilon + \eta_0 + \zeta_0 + \zeta_0\eta_0$; i.e., ϵ' has the form (20). Note in particular that $\epsilon\epsilon'\epsilon = \epsilon$.

Next, we have

$$\begin{aligned} \epsilon T = \epsilon'\epsilon\epsilon' &= (\epsilon + \zeta_0)(\epsilon + \eta_0)\epsilon(\epsilon + \zeta_0)(\epsilon + \eta_0) \\ &= (\epsilon + \zeta_0)(\epsilon + \eta_0) = \epsilon'. \end{aligned}$$

Clearly T is a linear transformation and a direct computation shows that $(\alpha_1 \alpha_2)T = (\alpha_1 T)(\alpha_2 T)$; i.e., T is an endomorphism on A.

To complete the proof that T is an automorphism, i.e. that it is one-to-one and onto we construct its inverse. Let $\alpha = \beta' + \eta' + \zeta' + \tau$ be the splitting (3) for α relative to ϵ' and let $\alpha T' = \epsilon\beta'\epsilon + \epsilon\eta' + \zeta'\epsilon + \tau$. Then the equations $TT' = T'T = I$ follow from $\epsilon\epsilon'\epsilon = \epsilon$ and $\epsilon'\epsilon\epsilon' = \epsilon'$.

The automorphism T of the theorem is completely defined by ϵ' and hence by η_0 and ζ_0; we denote it by $T(\eta_0, \zeta_0)$. It is easy to verify that the set W of all $T(\eta, \zeta)$ is a commutative group with composition rule

$$T(\eta_1, \zeta_1) \cdot T(\eta_2, \zeta_2) = T(\eta_1 + \eta_2, \zeta_1 + \zeta_2). \tag{21}$$

Let U_ϵ denote the subgroup consisting of all automorphisms of A which leave ϵ fixed; then the group G of all automorphisms has the factorization $U_\epsilon W$.

Let γ, γ' be elements of B for which $\gamma\gamma' = \gamma'\gamma = \epsilon$. Then the mapping $S(\gamma)$ given by

$$\alpha = \beta + \eta + \zeta + \tau \to \alpha S(\gamma) = \gamma'\beta\gamma + \gamma'\eta + \zeta\gamma + \tau \tag{22}$$

is an automorphism of A. The set V of all $S(\gamma)$ is an subgroup of U. We observe that N^2 is elementwise fixed under the automorphisms in V and in W.

According to Theorem 6, the general question of conditions for isomorphism of algebras of class Q can be reduced to the study of conjugacy (under automorphisms) of ideals contained in the square of the radical of a direct sum of submatrix algebras.

In this paper we shall limit our study of isomorphism to the case of algebras of class Q' and in particular those for which $K = F$, i.e. for which B is a total matrix algebra. According to Theorem 5 we can reduce this to the following question. Let $C = C(F, m, l, r)$, and let M, M' be two ideals in the square of the radical of C. We ask for necessary and sufficient conditions for the conjugacy of M and M' under automorphisms of C. We first determine the group G of automorphisms of C which leave F elementwise fixed.

We have initially the factorization $G = U_\epsilon W$. Let U_B denote the subgroup of G consisting of automorphisms which leave B elementwise fixed and let T be any element of U_ϵ. Then since B is a total matrix algebra T must agree on B with an inner automorphism of B, i.e., there exists γ in B such that $T(\gamma)^{-1}$ leaves B elementwise fixed **(1)**. This shows that $U_\epsilon = U_B V$. Now since W and V both leave N^2 elementwise fixed, M and M' will be conjugate under G if and only if they are conjugate under U_B.

THEOREM 8. *Let G be the group of automorphisms of a submatrix F-algebra $C = C(F, m, l, r)$. Then G has the factorization*

$$G = U_B V W$$

where W is isomorphic to a vector space of dimension m^2lr over F, where V is isomorphic to the full linear group $GL(m)$, and where U_B is isomorphic to the direct product of $GL(l)$ and $GL(r)$.

The statement about W follows from (21) and the fact that for C we have dim $\epsilon N = lm$ and dim $N\epsilon = rm$. The statement about V follows from (22) since $GL(m)$ is the group of inner automorphisms of a total matrix algebra of degree m.

Next, let T be any element of U_B, and choose e as in the proof of Theorem 5. Then, since e is in B, $eT = e$ and hence $(eN)T = eN$ and $(Ne)T = Ne$; moreover, T induces non-singular linear transformations T_L on eN and T_R on Ne. Let τ be an element of N^2. Then we have (cf. (16))

$$\tau = \sum_{i,s} a_{is}\zeta_i\eta_s. \tag{23}$$

(The factor κ_h appearing in (16) does not appear here since $K = F$.) Now,

$$\tau T = \sum_{i,s} a_{is}(\zeta_i T_R)(\eta_s T_L); \tag{24}$$

hence T is completely determined by T_L and T_R. It follows that the mapping $T \to (T_L, T_R)$ is a homomorphism of U_B into $GL(l) \times GL(r)$. Moreover, it follows from (24) that if $T \to (I_l, I_r)$, then $T = I$, i.e. this mapping is an isomorphism. Finally we show that the mapping is onto. Let T_L, T_R be any elements of $GL(l)$ $GL(r)$ respectively. Then the mapping $T = T(T_L, T_R)$ defined by

$$(\beta + \eta + \zeta + \tau)T = \beta + \eta T_L + \zeta T_R + \tau', \tag{25}$$

where τ' is given by (24), is clearly an element of U_B which maps into (T_L, T_R).

Now let M be an ideal of C contained in N^2, and let M have dimension g over F. Then (cf. (17)) M has a basis of the form

$$\{\tau_q = \sum_{i,s} a_q(i,s)\zeta_i\eta_s \ (q = 1, \ldots, g)\} \tag{26}$$

where the $a_q(i,s)$ are elements of F. We can associate M and the given basis with the trilinear form

$$f_M: f_M(x, y, z) = \sum_{q,i,s} a_q(i,s)\, x_q y_s z_i.$$

A change of basis for M replaces f_M according to a non-singular linear transformation on the x_q. Under an automorphism $T(T_L, T_R)$, M is replaced by a new ideal M' whose corresponding trilinear form is obtained from f_M by applying the substitutions T_L to the y_s and T_R to the z_i. Thus we see that *two ideals M and M' are conjugate under U_B and therefore under G if and only if their corresponding trilinear forms f and f' are equivalent.* Thus the problem of isomorphism of two algebras of class Q' (having $K = F$) is reduced to the equivalence of trilinear forms.

To extend this result to the case where $K \neq F$ would involve equivalence of quadrilinear forms under the full linear group on three of the sets of variables and under a finite group corresponding to automorphisms of K over F on the fourth set of variables. If the centre of K is inseparable over F, then Theorem 8 still remains valid except that U_B must be enlarged to account

for automorphisms of K. The factorization $G = U_B VW$ is no longer valid (unless the centre of K is inseparable over F). We leave the detailed analysis of this case as well as the general problem of isomorphism of algebras of class Q for future treatment.

References

1. E. Artin, C. J. Nesbitt, and R. M. Thrall, *Rings with minimum condition* (University of Michigan Press, 1944).
2. W. P. Brown, *Generalized matrix algebras*, Can. J. Math., 7 (1955), 188–190.
3. J. P. Jans, *On the indecomposable representations of algebras*. 2200–5–T Engineering Research Institute, University of Michigan.

University of Michigan

ON THE MODULAR REPRESENTATION OF THE SYMMETRIC GROUP
PART V

G. DE B. ROBINSON

1. Introduction. It has been observed **(2)** that the number of p-regular classes of S_n, i.e. the number of classes of order prime to p, is equal to the number of partitions (λ) of n in which no summand is repeated p or more times. For this relation to hold *it is essential that p be prime.* It seems natural to call the Young diagram $[\lambda]$ associated with (λ) *p-regular* if no p of its rows are of equal length, otherwise *p-singular.*

The problem considered here is that of refining the above result to prove (§5) that the number of regular diagrams in a given block is equal **(1; 2; 4; 5; 6)** to the number of modular irreducible representations (indecomposables of the regular representation of S_n) in that block. It is interesting to see that all our machinery is required. For example, the notion of an r-Boolean Algebra associated with a given diagram **(8)**, which seemed somewhat of a curiosity at first, plays a central role. In particular, *complementation* in such an rBA can be interpreted in terms of the core and so has significance for the block as a whole (§§2, 3). Similarly, the construction of a diagram with a given core from a knowledge of its p-quotient (star diagram) **(3, 7)** has to be made explicit (§4). This shows up the underlying number-theoretic basis of the theory in a new and significant light.

Actually, we are laying the foundation here for the establishment of an explicit correspondence between the indecomposables of the regular representation and the modular irreducible representations of S_n. The existence of this correspondence for any finite group was demonstrated by R. Brauer and C. J. Nesbitt in 1937, using very general arguments.

2. The complement of a Young diagram in its r-Boolean Algebra. Consider a Young diagram $[\lambda]$ to (from) which can be added (removed) $d(d^*)$ nodes of class r. Such a diagram belongs to an r-Boolean Algebra **(8)** of dimension $d + d^*$ in which the *complement* of $[\lambda]$ is obtained by removing the d^* r-nodes and adding r-nodes in the d free r-positions of $[\lambda]$. Let us denote this uniquely defined complement by $[\bar{\lambda}]$. As was shown in **(8)**

$$d - d^* = \delta, \tag{2.1}$$

so that the core $[\bar{\lambda}_0]$ of $[\bar{\lambda}]$ is obtainable by adding δ r-nodes to $[\lambda_0]$ where δ is the r-defect of $[\lambda_0]$. Also, the *weight* of $[\bar{\lambda}]$ is the same as that of $[\lambda]$. Two

Received December 26, 1954.

diagrams of the same block may belong to different rBA's and d and d^* may be different, but δ *is the same in each case.* We prove the following theorem (cf. **(8**, 4.4**)**):

2.2 *The p-quotients of* $[\lambda]$ *and its complement* $[\bar{\lambda}]$ *in the appropriate* rBA *are the same except for the interchange of the* r *and* $(r-1)$*-constituents.*

Proof. We distinguish three cases.

(i) If the ith row of $[\lambda]$ ends in a removable r-node and the jth column in an r-position **(8)** then $h_{ij} \equiv 0 \pmod{p}$. Clearly, removing the r-node at the end of the ith row and adding an r-node at the foot of the jth column does not change the (i,j)-hook length so that $\bar{h}_{ij} = h_{ij} \equiv 0 \pmod{p}$. It does however change the residue class of the hook from r to $r-1$. Similarly, if the ith row ends in an r-position and the jth column in a removable r-node the length of the (i,j) hook remains the same but its class is changed from $r-1$ to r.

(ii) If the ith row of $[\lambda]$ ends in a removable r-node and the jth column in an $(r+1)$-node below which *no* r-node can be added as in Figure 1,

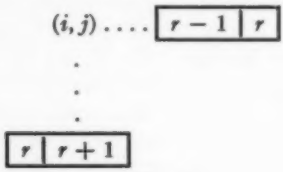

FIG. 1

then removing the r-node yields $\bar{h}_{i,j-1} = h_{i,j} \equiv 0 \pmod{p}$. Similarly, if the th column ends in a removable r-node and *no* r-node can be added at the end of the ith row as in Figure 2,

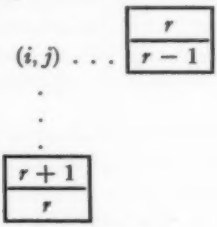

FIG. 2

then removing this r-node yields $\bar{h}_{i-1,j} = h_{i,j} \equiv 0 \pmod{p}$. Conversely, we may think of *adding* an r-node in the appropriate r-positions in Figures 1 and 2 to yield similar conclusions.

(iii) In this third case we have to consider (i,j)-hooks such that *no* r-node can be added or removed from the end of the ith row or the jth column. The situation is as indicated in Figure 3.

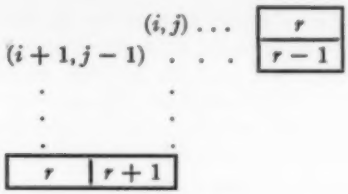

FIG. 3

Taken in conjunction with case (ii) it is clear that if the r-constituent of $[\lambda]_p$ receives a contribution from $h_{ij} \equiv 0 \pmod{p}$ in case (iii), then also the $(r-1)$-constituent of $[\bar{\lambda}]_p$ will receive a contribution from $h_{i+1,j-1} \equiv 0 \pmod{p}$ and *vice versa*. All three cases are concordant, in that an $h \equiv 0 \pmod{p}$ remains fixed in case (i), or moves one place in its row or column in case (ii), or diagonally in case (iii), so that the r and $r-1$ constituents of $[\lambda]_p$ and $[\bar{\lambda}]_p$ are interchanged, the others remaining unaltered.

2.3 *Example.* If $[\lambda] = [7, 5, 4^3, 2^2, 1^2]$ for $p = 3, r = 1$, then $d = 2, d^* = 2$ so that

$$[\lambda]_3 = [3^3], [1], —,$$

where the constituents of $[\lambda]_3$ are associated with the residue classes 0, 1, 2 respectively. We have $[\bar{\lambda}] = [8, 6, 4^3, 2, 1^2]$ and

$$[\bar{\lambda}]_3 = [1], [3^3], —.$$

Each of the changes described in cases (i), (ii), (iii) is illustrated.

3. Regular Young diagrams. Consider the 0-element $[\lambda^0]$ of an rBA for which $d^* = 0$ and let us think of adding an r-node in each of the possible $d = \delta$ r-positions **(8)**, by 2.1. It is possible that the addition of an r-node will lead to a singular diagram; if so, we shall call the corresponding r-position a *singular* position.

Corresponding to a given singular position P there exists a *regular* position P' at which an r-node can be added as indicated in Figure 4, the resulting diagram being regular.

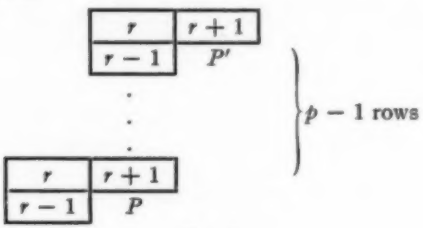

FIG. 4

Of course P' may itself be a singular position as in the Example 3.1 (whose corresponding regular position is $P'' = (3, 3)$). But the sequence of singular positions must eventually yield a regular position, since adding an r-node in the first row of $[\lambda]$ cannot yield a singular diagram.

Clearly, adding all δ r-nodes to $[\lambda^0]$ yields its complement $[\lambda^l]$, and if $[\lambda^0]$ is regular then so also is $[\lambda^l]$, since all singular positions and their corresponding regular positions are filled. A similar argument shows that $[\lambda^l]$ is singular if $[\lambda^0]$ is singular, and conversely.

When we consider an arbitrary element of an rBA for which $0 < d^* < d + d^* = l$ the situation is somewhat different, since if a regular position is occupied by an r-node, then in the complement the corresponding singular position will be occupied and the diagram will be singular. To overcome this difficulty we define a *modified complement* This definition favours regular diagrams, but a similar one would favour singular diagrams.

DEFINITION. If in a regular diagram $[\lambda]$, a singular r-position is vacant while its corresponding regular position is occupied by an r-node, then the *modified* complement of $[\lambda]$ in the appropriate rBA is that diagram obtained from $[\bar{\lambda}]$ by raising all r-nodes which occupy singular positions to the corresponding regular positions. Clearly, the modified complement of a modfiied complement is the original diagram.

3.1 *Example.* If $[\lambda] = [7, 4, 3, 2, 1^2]$, then in the complement for $p = 3$, $r = 0$, $[\bar{\lambda}] = [6, 5, 2^2, 1^2]$ the two singular positions $(5, 2)$ and $(7, 1)$ are occupied and the regular position $(3, 3)$ is vacant. Thus the modified complement is obtained by raising these 0-nodes to the corresponding regular positions to yield $[6, 5, 3, 2^2, 1]$.

We may think of complementation in the ordinary sense as a special case of modified complementation, and so state the following theorem:

3.2 *The property that a diagram be regular is invariant under modified complementation in the appropriate rBA.*

We conclude that (cf. **(7**, 5.6**)**):

3.3 *The number of p-regular diagrams of a given weight w is independent of the core.*

Proof. Complementation or modified complementation amounts to adding δ r-nodes to the core. But we know that the result is again a core, and every core can be obtained in this way **(8)**.

4. The explicit construction of a Young diagram with given p-core and p-quotient. Our construction is explicit and merely reverses an argument given elsewhere **(3)**.

Let us call the set of first column hook lengths obtained from $[\lambda_0]$ a *core set*, denoting them

4.1 $$\Gamma_k : c_1, c_2, \ldots, c_k, \qquad c_i > c_{i+1}.$$

If these c's are divided into residue classes, then it is known (8; 9) that the zero class is empty and that all residues smaller than the largest in a given class necessarily appear.

It is in general necessary to *extend* Γ_k in the following manner:

$$4.2\ \Gamma_h : b_1 = c_1 + s,\ b_2 = c_2 + s, \ldots b_k = c_k + s,\ b_{k+1} = s-1, \ldots b_{h-1} = 1,\ b_h = 0,$$

where $k + s = h$; we shall call Γ_h a *basic set*. Again we may divide the elements of Γ_h into residue classes, the zero class now appearing for $h > k$.

We shall denote the p-quotient $[\lambda]_p$ by the set of disjoint diagrams

$$4.3 \qquad [{}_0\lambda],\ [{}_1\lambda], \ldots, [{}_{p-1}\lambda],$$

where one or more constituents may be vacuous. The partition corresponding to a given constituent may be written out in detail thus:

$$4.4 \qquad [{}_r\lambda] = [{}_r\lambda_1, {}_r\lambda_2, \ldots, {}_r\lambda_{k_r}],$$

where r designates the residue class of the constituent.

In the required construction of a diagram $[\lambda]$ we must extend the core set Γ_k so that the quantities

$$4.5 \qquad {}_r\lambda_t \cdot p \qquad (t = 1, 2, \ldots, k_r)$$

can be added in order to the k_r largest members in the appropriate residue class of Γ_h. It only remains to determine this residue class for all r, and we do this by setting

$$4.6 \qquad r \equiv b_t - h \pmod{p}.$$

If we denote by g_r the number of elements in a core set which are congruent to $r \pmod{p}$, then the number of elements in the basic set Γ_h which are congruent to $r \pmod{p}$ is given by

$$4.7 \qquad g(r, s) = g_{r-s} + \left[\frac{s + p - 1 - r}{p}\right],$$

where the bracket function $[x]$ denotes the largest integer equal to or less than x. For a given $[\lambda]_p$ we have a set of integers $k_{r'} (r' = 0, 1, \ldots, p-1)$, and the choice of s for a given core is determined by the following conditions:

$$4.8 \qquad g(r, s)\begin{cases} > k_{r'}, & r' \not\equiv -h \pmod{p}, \\ = k_{r'}, & r' \equiv -h \pmod{p}, \end{cases}$$

where $r' \equiv r - h \pmod{p}$ by 4.6. That these conditions determine s *uniquely* (7) will be illustrated by the following

4.9 *Example.* To construct $[\lambda]$ given $[\lambda_0] = [2, 1^2]$, $[\lambda]_3 = [2, 1]$, $[1^2]$, $[2]$ where the constituents of $[\lambda]_3$ are associated with the residue classes 0, 1, 2 respectively. Table I gives the values of the function $g(r, s)$ and is arranged according to the residue classes $r' \equiv r - h \pmod{p}$, for $p = 3$.

Table I

s	h	$g(r', s)$		
		$r' = 0$	$r' = 2$	$r' = 1$
0	3	**0**	1	2
1	4	0	**2**	2
2	5	0	2	**3**
3	6	**1**	2	3
4	7	1	**3**	3
5	8	1	3	**4**
6	9	**2**	3	4
7	10	2	**4**	4
8	11	2	4	**5**

The integers printed in bold type are those which correspond to the *equality* in 4.8. We can use the table to determine s. Clearly the k's of $[\lambda]_3$ are $k_0 = 2$, $k_2 = 1$, $k_1 = 2$, so that $s \geqslant 6$. That $s \not> 6$ follows from the equality part of 4.8. Thus the basic set is

$$10, 8, 7, 5, 4, 3, 2, 1, 0.$$

Corresponding to $r' = 0$ we must add 6 to 3 and 3 to 0; corresponding to $r' = 1$ we must add 3 to 10 and 3 to 7; corresponding to $r' = 2$ we must add 6 to 8. Rearranging, we have the set of first column hook lengths:

$$14, 13, 10, 9, 5, 4, 3, 2, 1,$$

which belongs to $[6^2, 4^2, 1^3]$.

5. The enumeration of p-regular diagrams. As we have remarked, a core set can have no element $\equiv 0 \pmod{p}$, and moreover every class of elements must contain every integer less than the largest in the class. If a diagram is singular then there will be at least p *successive* integers in the set of first column hook lengths. Conversely this condition is also sufficient for singularity.

5.1 *Any diagram obtained by adding p-hooks to a p-core, while not increasing the number of rows, is necessarily regular.*

Proof. By adding multiples of p to a core set we can never introduce the zero residue class and so have p consecutive first column hook lengths.

The enumeration problem of regular diagrams will be solved by showing how we can enumerate singular diagrams of a given weight w, *provided the core is suitably chosen.* Having established the enumeration in one case then, by 3.3, it applies in all cases. The choice of the core affects the situation in a somewhat subtle manner as we can see by examining first the case $p = 2$. Here

there is only one type of core and by taking the number of rows to be g we have the corresponding core set to be

$$\Gamma_g : 2g-1, 2g-3, \ldots, 3, 1.$$

If we extend Γ_g to Γ_{g+s} where s is *even* it will always be possible to obtain a set of at least two consecutive terms provided $w \leqslant g+1$, and the diagrams have non-vacuous $[,\lambda]$, where $r \equiv -g \pmod 2$. If $w > g+1$ then the diagram represented by

$$[,\lambda] = [w],$$

$r \equiv -g \pmod 2$, will have no two consecutive terms and so will be regular.

Moreover, if s is *odd* all the terms in the extended set are even except some of those at the end so it will be necessary to add (from the top down) at least $g+2$ multiples of 2 to obtain a set of first column hook lengths, i.e. $w > g+1$.

By denying singularity we obtain regularity so we have proved the following theorem:

5.3 *For $p = 2$, and $w \leqslant g+1$ the necessary and sufficient condition that a diagram be regular is that in its 2-quotient the constituent $[,\lambda]$ be vacuous for $-r \equiv h \equiv g \pmod 2$.*

For $p > 2$ the situation is complicated by the fact that there are $p-1$ classes of terms in a core set and some of these may be vacuous. Thus to produce a set of p consecutive first column hook lengths in the manner envisaged above the weight w will be limited by the *shortest* residue class in the core set; we call the length g of this shortest class the *grade* of the core. If any class is vacuous, $g = 0$, and $w = 1$ is the only possible case yielding such an enumeration of singular diagrams.

The following theorem includes 5.3 as a special case.

5.4 *If a p-core set Γ_k contains $p-1$ residue classes, each class containing at least g members, then the necessary and sufficient condition that a diagram $[\lambda_1, \lambda_2, \ldots, \lambda_k]$ of weight $w \leqslant g+1$ be regular is that $[,\lambda]$ be vacuous for $-r \equiv h \equiv k \pmod p$.*

Proof. As before we fix attention on the singular diagrams. From our definition of the grade g, we know that there are at most g sets of $p-1$ consecutive residues in Γ_k. The worst case we need consider is where $[,\lambda] = [w]$ so that $s = p$. Adding wp to zero we obtain Γ_h which contains the consecutive set

$$wp, wp-1, \ldots, (w-1)p+1,$$

so that $[\lambda]$ is singular. The other extreme case is where $[,\lambda] = [1^w]$ and we add p to each of w terms which are all congruent to zero (mod p). In this case we have the consecutive terms

$$wp, wp-1, \ldots, 3, 2, 1,$$

and $[\lambda]$ is certainly singular. All other partitions of w clearly yield at least p consecutive terms in Γ_k.

If the w nodes are distributed over more than one constituent of $[\lambda]_p$ then, provided $[,\lambda]$ is not vacuous, the above argument is still applicable and $[\lambda]$ must be singular.

On the other hand if $[,\lambda]$ is vacuous, $h \equiv k \pmod p$, and since additions must be made at the top of a residue class of Γ_h, $w > g + 1$ as in 5.3, proving that all diagrams under consideration must be regular.

Denoting the number of partition of w, by $\mathfrak{p}_w$, we combine 3.3 and 5.4 and conclude **(1; 2; 4; 5; 6)** that:

5.5 *The number of p-regular diagrams of weight w having a given core $[\lambda_0]$ is equal to*

$$\sum_{w_1 \ldots w_{p-1}} \mathfrak{p}_{w_1} \mathfrak{p}_{w_2} \cdots \mathfrak{p}_{w_{p-1}} \qquad \left(\sum_1^{p-1} w_i = w, 0 \leqslant w_i \leqslant w\right)$$

and so equal to the number of modular irreducible representations of S_n in the corresponding block.

We prove in conclusion the following interesting result:

5.6 *For $w < g + 1$ the diagrams in a given block are all zero elements $(d^* = 0)$ of their respective rBA's, where $-r \equiv k \pmod p$.*

Proof. If $d^* \neq 0$ there is a removable node of class r; this implies the presence of a term of Γ_h congruent to zero (mod p), *followed* by a gap in the set. Since this does not happen for $w < g + 1$, with $-r \equiv k \pmod p$, we conclude that $d^* = 0$ for every such diagram.

Consider a core C_1 of grade g having k rows. Clearly, the next succeeding node in the first column would be of class r where $-r \equiv k \pmod p$. If δ such r-nodes are added to C_1 we obtain a core C_2. Successive complementation in this manner yields a *series* of cores. For $p = 3$,

(a) [2], [3, 1], [3, 1^2], [4, 2, 1^2], [4, 2^2, 1^2], . . . ,
(b) zero, [1], [1^2], [2, 1^2], [2^2, 1^2], [3, 2^2, 1^2], . . . ,

are two such series. We state the following lemmas without proofs.

5.7 *Two distinct series cannot have a core in common.*

5.8 *The grade cannot decrease under complementation.*

It follows from 5.6 that if $w < g+ 1$ for diagrams having C_1 as core, then subsequent complementation with $-r \equiv k \pmod p$ is *ordinary* and the diagrams so obtained have C_2, C_3, etc. as cores. Moreover, regular diagrams go into regular diagrams and singular into singular by 2.2, 5.4 and 5.8. Clearly, the critical class of the p-quotient which is vacuous for regular diagrams by 5.4 is permuted according to the cycle

$$(0, p - 1, p - 2, \ldots 2, 1),$$

under successive complementation in the series.

5.9 *Example.* To illustrate these ideas we give in Table II the sets Γ_h for 3-singular diagrams with core $[2^2, 1^2]$ for which $\Gamma_k = 5, 4, 2, 1$ and $g = 2$, of weight $w = 3$, and the associated sets of first column hook lengths. The arrangement of these should be monotonic to construct $[\lambda]$, but they are left as they come after the appropriate multiples of p are added to the terms of Γ_h.

Table II

$[\lambda]_3$	Γ_h	First column hook lengths	$[\lambda]$
–, –, [3]	8, 7, 5, 4, 2, 1, 0	8, 7, 5, 4, 2, 1, **9**	$[3^3, 2^2, 1^2]$
–, –, [2,1]	11, 10, 8, 7, 5, 4, 3, 2, 1, 0	11, 10, 8, 7, 5, 4, **9**, 2, 1, **3**	$[2^6, 1^5]$
–, –, $[1^3]$	14, 13, 11, 10, 8, 7, 6, 5, 4, 3 2, 1, 0	14, 13, 11, 10, 8, 7, **9**, 5, 4, **6**, 2, 1, **3**	$[2^3, 1^{11}]$
[1], –, [2]	8, 7, 5, 4, 2, 1, 0	8, **10**, 5, 4, 2, 1, **0**	$[4, 3, 2^2, 1^2]$
[1], –, $[1^3]$	11, 10, 8, 7, 5, 4, 3, 2, 1, 0	11, **13**, 8, 7, 5, 4, **6**, 2, 1, **3**	$[4, 3, 1^5]$
[2], –, [1]	8, 7, 5, 4, 2, 1, 0	8, **13**, 5, 4, 2, 1, **3**	$[7, 3, 1^5]$
$[1^2]$, –, [1]	8, 7, 5, 4, 2, 1, 0	8, **10**, 5, 7, 2, 1, **3**	$[4, 3^2, 2, 1^3]$
–, [1], [2]	8, 7, 5, 4, 2, 1, 0	**11**, 7, 5, 4, 2, 1, **6**	$[5, 2^4, 1^2]$
–, [1], $[1^2]$	11, 10, 8, 7, 5, 4, 3, 2, 1, 0	**14**, 10, 8, 7, 5, 4, **6** 2, 1, **3**	$[5, 2, 1^5]$
–, [2], [1]	8, 7, 5, 4, 2, 1, 0	**14**, 7, 5, 4, 2, 1, **3**	$[8, 2, 1^5]$
–, $[1^2]$, [1]	8, 7, 5, 4, 2, 1, 0	**11**, 7, **8**, 4, 2, 1, **3**	$[5, 3^2, 1^4]$
[1], [1], [1]	8, 7, 5, 4, 2, 1, 0	**11**, **10**, 5, 4, 2, 1, **3**	$[5^2, 1^5]$

References

1. H. Farahat, *On the representations of the symmetric group*, Proc. London Math. Soc. (3), *4* (1954), 303–316.
2. J. S. Frame and G. de B. Robinson, *On a theorem of Osima and Nagao*, Can. J. Math., *6* (1954), 125–127.
3. J. S. Frame, G. de B. Robinson and R. M. Thrall, *The hook graphs of the symmetric group*, Can. J. Math., *6* (1954), 316–324.
4. H. Nagao, *Note on the modular representations of symmetric groups*, Can. J. Math., *5* (1953), 356–363.
5. M. Osima, *Some remarks on the characters of the symmetric group*, Can. J. Math., *5* (1953), 336–343.
6. ———, II, *ibid.*, *6* (1954), 511–521.
7. G. de B. Robinson, *On a conjecture by J. H. Chung*, Can. J. Math., *4* (1952), 373–380.
8. ———, *On the modular representations of the symmetric group* IV, *ibid.*, *6* (1954), 486–497.
9. R. M. Thrall and G. de B. Robinson, *Supplement to a paper by G. de B. Robinson*, Amer. J. Math., *73* (1951), 721–724.

University of Toronto

PSEUDO-REGULARITY

NATHAN DIVINSKY

Introduction. An element x is said to be *right-quasi-regular* (r.q.r.) if there exists an element y such that $x + y + xy = 0$. This concept had its inception in the fact that (for rings with unity) if $1 + x$ has an inverse, written as $1 + y$, then $(1 + x)(1 + y) = 1, x + y + xy = 0$. Thus in rings without unity elements it seemed **(1; 3; 12)** profitable to consider this latter equation. Jacobson **(9)** was able to employ this concept in obtaining a structure theory for rings without chain conditions.

Our point of departure is in considering the expression $x + y + xy$ not as stemming from $(1 + x)(1 + y)$, but as a special case of the more general expression $x + x^n y + x^{n+1}y$. Our considerations seem to bear most fruit in the case $n = 1$ for commutative rings. We call an element x *right-pseudo-regular* (r.p.r.) if there exists an element y such that $x + xy + x^2y = 0$.

In §1 we show the existence of a maximal r.p.r. ideal R, called the *subradical*; and show that with some mild restrictions on the ring, it is simply the Jacobson radical J, thus obtaining a new representation of J. In general however, $R \leqslant J$. We also obtain some radical-like properties of R, as well as a definite relationship between R and J.

In §2 we use the techniques of Brown and McCoy and in the commutative case are able to show that $A - R$ is isomorphic to a subdirect sum of subdirectly irreducible rings, some of which are simple with unity (fields) and others are bound to their maximal nil ideal in the sense of Hall **(8)**.

1. An element x of a ring A shall be called *right-pseudo-regular* (r.p.r.) of degree n, if there exists an element y of A such that

$$x + x^n y + x^{n+1}y = 0.$$

It is clear that for $n = 0$ we get the familiar right-quasi-regularity. We shall be primarily interested in the case $n = 1$, and refer to it simply as r.p.r. It is also clear that if x is r.p.r. of degree n, then x is r.p.r. of degree $n - 1$

$$x + x^{n-1} \cdot xy + x^n \cdot xy = 0;$$

and so in particular, if x is r.p.r. it is r.q.r. (*right-quasi-regular*). The converse of this last statement is not true, since in the ring of even integers modulo 4, the element 2 is r.q.r., since

$$2 + 2 + 2.2 \equiv 0 (\mathrm{mod}\ 4),$$

Received January 5, 1955. This paper was written while the author attended the Summer Research Institute of the Canadian Mathematical Congress, 1953. It was presented to the American Mathematical Society on November 27, 1953.

but it is not r.p.r. The exact relationship between right-pseudo-regularity and right-quasi-regularity is obtained in

LEMMA 1. *An element x of a ring A is* r.p.r. *of degree n, $x + x^n y + x^{n+1} y = 0$, if and only if x is* r.q.r. *and there exists an element x' such that $x^n x' = x$.*

Proof. If x is r.p.r. of degree n, then clearly x is r.q.r. and $x^n x' = x$ with $x' = -y - xy$. Conversely if x is r.q.r., $x + z + xz = 0$, and if there exists an x' such that $x^n x' = x$, then setting $y = -x' - x'z$ we find that

$$\begin{aligned} x + x^n y + x^{n+1} y &= x + x^n(-x' - x'z) + x^{n+1}(-x' - x'z) \\ &= x - x - xz - x^2 - x^2 z \\ &= -x(z + x + xz) = 0. \end{aligned}$$

COROLLARY 1. *An element x of a ring A is* r.p.r. *if and only if x is* r.q.r. *and there exists an element x' in A such that $xx' = x$.*

COROLLARY 2. *If x is in xA for every x of A, then right-quasi-regularity and right-pseudo-regularity are equivalent concepts.*

A more unexpected result is

LEMMA 2. *Right-pseudo-regularity of degree 2 and right-pseudo-regularity of degree n for all $n > 1$, are equivalent concepts.*

Proof. Clearly, if x is r.p.r. of degree n, $n > 1$, it is r.p.r. of degree 2. Conversely, if x is r.p.r. of degree 2, then by Lemma 1, x is r.q.r. and there exists an x' such that $x^2 x' = x$. Notice that this is precisely strong regularity. Then

$$x^n \cdot x'^{n-1} = x^{n-2} \cdot x^2 x' \cdot x'^{n-2} = x^{n-1} \cdot x'^{n-2} = \ldots = x^2 x' = x.$$

Thus there is an element $w = x'^{n-1}$ such that $x^n w = x$. Therefore by Lemma 1, x is r.p.r. of degree n.

A right ideal Q will be called r.p.r. of degree n if all of its elements are r.p.r. of degree n. To consider the existence of maximal r.p.r. of degree n right ideals we shall make use of the following

LEMMA 3. *If $x \neq 0$, is* r.p.r., *$x + xy + x^2 y = 0$, then its right-pseudo-inverse* (r.p.i.) *y is not in the Jacobson radical J.*

Proof. If y is in J, then $y + xy$ is in J and there exists an element z such that

$$y + xy + z + (y + xy)z = 0.$$

We have

$$\begin{aligned} 0 &= x + xy + x^2 y + (x + xy + x^2 y)z \\ &= x + x(y + xy + z + yz + xyz) \\ &= x + x \cdot 0 = x. \end{aligned}$$

This contradicts the fact that $x \neq 0$ and therefore y is not in J.

If we can show the existence of a maximal right ideal R_n, which is r.p.r. of degree n, for every n, then by Lemma 2, $R_2 = R_3 = \ldots = R_n$. This is in fact true but they are all equal to zero!

THEOREM 1. *If, for any $n > 1$, Q is a right ideal all of whose elements are* r.p.r. *of degree n, then $Q = 0$.*

Proof. Let Q be r.p.r. of degree 2. If x is in Q then $x + x^2y + x^2y = 0$. Then x is r.p.r. with r.p.i. xy. But xy is in Q; and since Q is a right ideal all of whose elements are r.q.r., Q is in J, and xy is in J. This contradicts Lemma 3 unless $x = 0$, $Q = 0$.

Though Theorem 1 proves that there are no right ideals all of whose elements are r.p.r. of degree n, $n > 1$, there may be many elements which are r.p.r. of degree n. Let A be a division ring. Then every element $\neq -1$ is r.q.r. Furthermore $x^2x^{-1} = x$ and thus by Lemma 1, every element $\neq -1$ is r.p.r. of degree 2 and thus by Lemma 2, every element $\neq -1$ is r.p.r. of degree n for every n.

We shall now show the existence of a maximal r.p.r. ideal, which is not always zero. The first step is to show that the sum of two r.p.r. right ideals is again an r.p.r. right ideal. To this end we prove a slightly more general result akin to Kaplansky's **(10**, Lemma 1**)**.

LEMMA 4. *If x is* r.p.r. *and if a belongs to an* r.p.r. *right ideal Q, then $x + a$ is* r.p.r.

Proof. By Lemma 1, Corollary 1, it is sufficient to show that $x + a$ is r.q.r. and that there exists an element v such that $(x + a)v = x + a$. The fact that $x + a$ is r.q.r. follows immediately from Kaplansky's lemma, since x being r.p.r. is also r.q.r. and Q being an r.p.r. right ideal is an r.q.r. right ideal.

To find the element v, we first define $u = a - ax'$ where $xx' = x$. Since a is in Q, u is in Q, and there exists an element u' such that $uu' = u$. Define $v = x' + u' - x'u'$. Then $(x + a)v = x + a$ follows from $xv = x$ and

$$av = ax' + (a - ax')u' = ax' + uu' = ax' + u = a.$$

We now define R to be the join of all the r.p.r. right ideals of the ring A. By Lemma 4, R is an r.p.r. right ideal. It is clear that R is the set of all elements that generate r.p.r. right ideals, i.e., all x such that $xi + xa$ is r.p.r. for every integer i and every element a of A. We now show that R is a two-sided ideal.

Let x be any element in R and a be any element in A. We must show that ax is in R, i.e., that $axi + axb$ is r.p.r. for every integer i and every b of A. Since x is in R, $xi + xb$ is in R. Let $y = xi + xb$. Then it is sufficient to show that ay is r.p.r. for any y in R. Since R is a right ideal, ya is in R and therefore there exists a z such that $ya + yaz + (ya)^2z = 0$. Then

$$ay + (-ay - ayayzy) + ay(-ay - ayayzy)$$
$$= ay - ay - a(yaz + ya + yayaz)y = 0 - a \,.\, 0 \,.\, y = 0.$$

Therefore ay is r.q.r. Furthermore, since y is in R, y is r.p.r., there exists a y' such that $yy' = y$. Therefore $ay \cdot y' = ay$. Therefore, by Lemma 1, Corollary 1, ay is r.p.r. We have proved

THEOREM 2. *If A is an arbitrary ring, the join R of all the* r.p.r. *right ideals of A is an* r.p.r. *two-sided ideal.*

We shall call R the right subradical of A. Most of the time $R < J$, but using Corollary 2, Lemma 1 we have

THEOREM 3. *If x is in xA for every x of A, or of J, then $J = R$.*

By considering left-pseudo-regularity we could, by exactly the same techniques, prove the existence of a maximal l.p.r. two-sided ideal L, which we call the left subradical of A. We would find that if A had a left unity, or if x was in Ax for every x of J, then $J = L$. It should be clear that though J enjoys certain left-right symmetric properties, the ideals R and L have no such well-roundedness. Of course if A has a unity element, then $J = R = L$; however in the general case, R and L are different. To see this, consider the set A of all two by two matrices of the form

$$\begin{pmatrix} a & 0 \\ b & 0 \end{pmatrix}$$

where a and b are integers mod 4. Then A contains 16 elements. The Jacobson radical J has 8 elements, namely those with $a = 0$ or 2, and $b = 0$, 1, 2 or 3. Furthermore, A has a right unity,

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

(in fact A has four different right unity elements) and therefore by Theorem 3, $J = R$. However A does not have a left unity and one can easily see that L has only the one element

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Though R is occasionally equal to J, it is often equal to 0. By Lemma 3, it is clear that if A is a radical ring, $A = J$, then $R = 0$. And also therefore, there is no nonzero ring which is equal to its right subradical.

The right subradical R has the usual radical-like properties.

THEOREM 4. *The difference ring $A - R$ is sub-semi-simple, that is, it has zero subradical.*

Proof. Let $\bar{R}$ be the subradical of $\bar{A} = A - R$. If $\bar{x}$ is in $\bar{R}$, then there exists an element $\bar{y}$ in $\bar{A}$ such that $\bar{x} + \bar{x}\bar{y} + \bar{x}^2\bar{y} = \bar{0}$, that is, $x + xy + x^2y$ is in R. Then there exists an element z in A such that

$$x + xy + x^2y + (x + xy + x^2y)z + (x + xy + x^2y)^2z = 0.$$

Rewriting this we have

$$x + (x + x^2)(y + z + yz + yxz + xyz + yxyz + yx^2yz) = 0.$$

Therefore x is r.p.r. Furthermore, since $\bar{x}$ is in $\bar{R}$, $\bar{x}\bar{i} + \bar{x}\bar{a}$ is in $\bar{R}$ for every integer i and every $\bar{a}$ in $\bar{A}$. As for x, we can show that $xi + xa$ is r.p.r. for every i and a, and therefore x is in R, $\bar{x} = \bar{0}$, $\bar{R} = \bar{0}$.

Jacobson has shown **(9)** that $J_n = J(A_n)$, where A_n is the set of all n by n matrices with elements in A, J_n is the set of all n by n matrices with elements in J, and $J(A_n)$ is the Jacobson radical of A_n. The corresponding result for subradicals is true and the proof is straightforward.

THEOREM 5. *The subradical $R(A_n)$ is equal to R_n.*

LEMMA 5. *The subradical $R = RA^n$ for every integer n.*

Proof. Since R is an ideal, $RA^n \leqslant R$. Conversely, if x is in R, there exists an element x' such that $x = xx'$, x is in RA. Therefore $R \leqslant RA$, $R \leqslant RA^n$, $R = RA^n$.

LEMMA 6. *The subradical $R \leqslant M_l$, the intersection of all the maximal left ideals.*

Proof. Jacobson has shown **(9)** that $J \cdot A \leqslant M_l$. Since $R \leqslant J \cdot A$, $R \leqslant M_l$.

We shall now obtain a more definite relationship between J and R.

LEMMA 7. *Let A be a non-nilpotent ring with the descending chain condition on right ideals, having all its idempotents in the centre. In particular A may be any commutative ring with* d.c.c. *Then if $A^n = A^{n+1}$, A^n has a unity element. In particular if $A = A^2$, A has a unity element.*

Proof. By d.c.c. **(2)** there exists an idempotent e such that

$$A = Ae + B$$

where B is the set of all $x - xe$ for x in A, and B is nilpotent. Furthermore $Ae \cdot B = B \cdot Ae = 0$ since e is assumed to be in the centre. Therefore $A^2 = (Ae)^2 + B^2$,

$$A^n = (Ae)^n + B^n.$$

Since $A > A^2 > \ldots > A^m > \ldots$ is a descending chain of right ideals, there exists an integer n such that $A^n = A^{n+1}$. It is clear that $B^n = B^{n+1}$, since if x is in B^n then $xe = 0$. And since $x = a + b$, with a in $(Ae)^{n+1}$ and b in B^{n+1}, we have $xe = 0 = ae + be = a + 0$. Therefore $a = 0$ and $x = b$ in B^{n+1}. Therefore $B^n \leqslant B^{n+1} \leqslant B^n$. Since B is a nilpotent ideal, $B^n = 0$, $A^n = (Ae)^n$. But $(Ae)^n = A^n e^n = A^n e$. Then $A^n = A^n e$, e is a unity element for A^n. It is clear that $e = e^n$ is in A^n.

LEMMA 8. *If B is an ideal of A, then the radical of the ring B, $J(B)$, is equal to $J \cap B$, where J is the Jacobson radical of A.*

This result is due to Perlis **(13)**.

LEMMA 9. *If B is an ideal of A, then $R(B) \leqslant R \cap B$, where R is the subradical of A.*

The proof uses Lemma 8 and is straightforward. Note that it is impossible to prove that $R(B) = R \cap B$, since if we take $B = J$, $R(B) = 0$, whereas $R \cap B = R \cap J = R$.

THEOREM 6. *If A is a ring with* d.c.c. *on right ideals, having all its idempotents in the centre, then $R = JA^{n-1}$, where n is the smallest integer such that $A^n = A^{n+1}$. When $n = 1$, $R = J$.*

Proof. If A is nilpotent, $A^n = 0$, $R = JA^{n-1} = 0$ by the remark just before Theorem 4. If A is not nilpotent, by d.c.c. there exists a least integer n such that $A^n = A^{n+1} \neq 0$. Then by Lemma 7, A^n has a unity element and by Theorem 3, $R(A^n) = J(A^n)$. By Lemma 9, $R \geqslant R(A^n)$. By Lemma 8, $J(A^n) = J \cap A^n$. Therefore

$$R \geqslant R(A^n) = J(A^n) = J \cap A^n \geqslant JA^{n-1}.$$

Conversely, by Lemma 5, $R = RA^{n-1}$. Thus $R \leqslant JA^{n-1}$. Therefore $R = JA^{n-1}$.

By similar techniques we can show that the left subradical L is contained in M_r, the intersection of all the maximal right ideals, and that $L = A^{n-1}J$.

Discussion of Theorem 6. Theorem 6 is not true without d.c.c., as the following example, mentioned to the author in a discussion with Professor Zassenhaus, proves. Let x_α be a basis for a commutative algebra, where the α's are real, $0 \lneqq \alpha \lneqq 1$. Define $x_\alpha x_\beta = x_{\alpha+\beta}$ if $\alpha + \beta < 1$, and equal to 0 if $\alpha + \beta \geqslant 1$. Then it is clear that every element is nilpotent. Thus $A = J$, and $R = 0$. However $x_\alpha = x_{\frac{1}{2}\alpha} x_{\frac{1}{2}\alpha}$, and therefore $A = A^2$. To be sure, A is nil, but not nilpotent.

Whether the theorem is true if A has d.c.c., but not the restriction that the idempotents lie in the centre, seems to be an open question. Since every ring with d.c.c. can be expressed **(7)** as $A = M + M^*$ where M is the maximal regular ideal, and M^* is bound to its radical in the sense of Hall **(8)**, and $MM^* = M^*M = 0$, the condition $A = A^2$ implies $M^* = M^{*2}$. Thus the first step seems to be to decide whether there exists a ring, say B, with d.c.c., bound to its radical, without a right or left unity element and such that $B = B^2$.

Another attack on this question can be made using a technique due to Baer **(4)**. Since we can write $A = Ae + B$, where B is the set of all $x - xe$ with x in A, and with B in J, we embed Ae in a maximal left ideal F (we assume if necessary a.c.c.). Then $A = (F, J)$. Since J is contained in every maximal *modular*[1] left ideal, if F were modular, $A = F$, a contradiction. Thus either Ae is already A, in which case A has a right unity, $R = J$, and we are well away to proving Theorem 6; or F is not modular. The condition $A = A^2$, together

[1]A left ideal L is called *modular* if there exists an element e in the ring A, such that $xe - x$ is in L for every x of A.

with d.c.c., imply that every maximal ideal is modular and that for every maximal left ideal, and in particular for F, $A - F$ is an irreducible A-module. Then $A - F$ is A-isomorphic to $A - Q$ where Q *is* a maximal modular left ideal of A. It is not clear though that F must be modular.

With regard to the following properties:

(a) A has a right unity element;

(b) For every x in A, x is in xA;

(c) $A = A^2$;

it is interesting to observe that (a) implies (b), and (b) implies (c). Lemma 7 proves that with d.c.c. and idempotents in the centre, (c) implies (a) and thus that the three conditions are equivalent. The above-mentioned example due to Zassenhaus shows that (c) does not imply (b) without d.c.c. To see that (b) does not imply (a) without d.c.c., consider the set of all infinite diagonal matrices, elements in a field, each matrix having only a finite number of nonzero entries.

2. Following Brown and McCoy **(5; 6)**, we associate with every element a in A, the ideal

$$R'(a) = \{ax - a^2x + \sum y_i az_i - \sum y_i a^2 z_i\}.$$

We call an element a, R'-regular if a is in $R'(a)$. We call an ideal I, R'-regular if every element of I is R'-regular. In this way we obtain the set R'' of all elements that generate R'-regular ideals. The set R'' is simply a special case of Brown and McCoy's F-radical. If a is an element of the subradical R, then $-a$ is also in R,

$$-a - ab + a^2b = 0, \quad a = a(-b) - a^2(-b),$$

a is R'-regular. Thus it is clear that $R \leqslant R''$. In the commutative case $R = R''$, though in general they are different. From **(5)** we have the following important results about R''.

THEOREM 7. *The set R'' is an ideal of A.*

THEOREM 8. $R''(A - R'') = 0$.

THEOREM 9. *The ring $A - R''$ is isomorphic to a subdirect sum of subdirectly irreducible rings each having their $R'' = 0$.*

THEOREM 10. *A subdirectly irreducible ring A, has its $R'' = 0$ if and only if there exists an element $e \neq 0$ in the minimal ideal K of A such that $R'(e) = 0$.*

THEOREM 11. *A has its $R'' = 0$ if and only if it is isomorphic to a subdirect sum of subdirectly irreducible rings each having their $R'' = 0$.*

Let A be a subdirectly irreducible ring with $R'' = 0$. Then by Theorem 10, there exists an element $e \neq 0$ in the minimal ideal K of A such that $R'(e) = 0$. Thus

$$\{ex - e^2x + \sum y_i ez_i - \sum y_i e^2 z_i\} = 0.$$

Therefore $ex = e^2x$ for every x in A. Thus $e^2 = e^3$, $e^3 = e^4$, $e^2 = e^4$. We will use

LEMMA 10. *If there exists a nonzero idempotent e', both in the centre and in the minimal ideal K of a subdirectly irreducible ring A, then A is simple with e' as unity.*

Proof. Consider the Peirce decomposition of A for e'. $A = A_1 + A_2$, where $A_1 = Ae'$, and A_2 is the set of all $x - xe'$ for x in A and is the set of all x such that $xe' = 0$. Since A_2 is an ideal which cannot contain e', and since e' is in every nonzero ideal, $A_2 = 0$. Therefore $A = A_1 = Ae'$. Since e' is in K, $A = K$, A is simple ($A^2 = Ae' \cdot Ae'$ contains e' and is therefore not zero), and has e' as unity element.

Therefore if $e^2 \neq 0$ and e is in the centre, in the subdirectly irreducible ring A with $R'' = 0$, K contains a nonzero idempotent in the centre, and A is a simple ring with unity. Otherwise $e^2 = 0$, and then $ex = 0$ for every x in A. Thus $eA = 0$, $JA = 0$. We have proved

THEOREM 12. *If a ring A has all its idempotents in the centre, then it has $R'' = 0$ if and only if A is isomorphic to a subdirect sum of subdirectly irreducible rings some of which are simple with unity, others (call them B_t) have the property that $K_t B_t = 0$, where K_t is the minimal ideal of B_t.*

With some mild conditions on A we can remove the nonsimple components.

THEOREM 13. *If a ring A has all its idempotents in the centre and either of the following properties:* (a) *if $xA < I$, then x is in I, for every ideal I;* (b) *every ideal M such that $A - M$ is subdirectly irreducible, is modular; then A has $R'' = 0$ if and only if A is isomorphic to a subdirect sum of simple rings with unity.*

Proof. Consider the components B_t of Theorem 12, that are not simple, i.e., the ones that satisfy $K_t B_t = 0$. For every i, there exists an ideal M_t in A such that $A - M_t \cong B_t$. If A satisfies (a), then $e_t A < M_t$ implies e_t is in M_t. Thus if $e_t B_t = 0$ in B_t, $e_t A < M_t$, e_t is in M_t, $e_t = 0$ in B_t. This contradicts the fact that $e_t \neq 0$ in B_t. Therefore there are no nonsimple components. Finally, if A satisfies (b), then $A - M_t$ has a unity element. Therefore $e_t B_t$ cannot be zero.

We now turn our attention to the commutative case. Here $R = R''$.

McCoy **(11)** made a study of all commutative subdirectly irreducible rings and considered them in two large classes: those having at least one element not a divisor of zero, and others all of whose elements are divisors of zero. We obtain more information about them in

THEOREM 14. *If A is a commutative subdirectly irreducible ring with subradical R zero, and with at least one element not a divisor of zero, then A is a field.*

If A is a commutative subdirectly irreducible ring all of whose elements are divisors of zero, then its subradical is zero and it is bound to its maximal nilideal N, and therefore also bound to its Jacobson radical J. Furthermore, if A has either d.c.c. *or* a.c.c., *A is nilpotent.*

Proof. By Theorem 12, a commutative subdirectly irreducible ring A with subradical R zero is either a simple ring with unity or $eA = 0, e \neq 0$. Therefore if A is fortunate enough to possess an element which is not a divisor of zero, $eA \neq 0$, A is simple with unity, A is a field.

From **(11)** we learn that if A is a commutative subdirectly irreducible ring all of whose elements are divisors of zero, then the minimal ideal $K = (e, 2e, \ldots pe = 0)$ and $eA = 0$. Therefore $KA = 0$ and by Theorem 12, $R = 0$. Furthermore, if $A \neq N$, the maximal nilideal, then let x be any element not in N. Since in the commutative case N is the set of *all* nilpotent elements, x is not nilpotent, the ideal $xA \neq 0$, and the ideal $x^2A \neq 0$. Since e is in every nonzero ideal, $e = x^2y$. Then

$$(xy)^2 = x^2y \cdot y = ey = 0,$$

since $eA = 0$. Therefore xy is nilpotent, xy is in N. Thus $x \cdot xy = e \neq 0$, $xN \neq 0$. Thus if $xN = 0$, x must be in N, i.e. A is bound to N. Then of course A is bound to J, since if $xJ = 0$, $xN = 0$, x is in $N < J$.

Assume now that A has d.c.c. If A is not nilpotent, and therefore not nil, there exists an element x which is not nilpotent. Consider the chain $xA > x^2A > \ldots$, and get an integer n such that $x^nA = x^{n+1}A$. Thus there is an element y such that $x^{n+1} = x^{n+1}y$. Since $x^{n+1}A$ is a nonzero ideal, e is in it, $e = x^{n+1}z$. Then

$$ey = 0 = x^{n+1}yz = x^{n+1}z = e \neq 0.$$

This contradiction shows that x must be nilpotent, A is nil and therefore nilpotent.

Assume now that A has a.c.c. If A is not nil, then again let x be any non-nilpotent element. Then all the ideals x^iA are nonzero and therefore each of them contains e. Therefore

$$e = xy_1 = x^2y_2 = \ldots = x^ny_n = \ldots.$$

Define V_i to be the set of all z that annihilate x^i. It is clear that the V_i are ideals and that $V_1 < V_2 < \ldots < V_i < \ldots$. Since

$$ex = 0 = x^iy_ix = x^{i+1}y_i,$$

y_i is in V_{i+1}, but is not in V_i. Therefore this is a properly ascending chain which does not stop after a finite number of steps. (Since x is not nilpotent, no V_i is equal to the whole ring.) This contradicts a.c.c. and therefore A is nil. By a result of Zassenhaus as yet unpublished, which states that in the presence of a.c.c. the maximal nil ideal is nilpotent, A is nilpotent.

Combining Theorems 12 and 14 we have our main result:

THEOREM 15. *If A is a commutative ring whose subradical R is zero, then A is isomorphic to a subdirect sum of subdirectly irreducible rings $A_1, A_2, \ldots, B_1, B_2, \ldots$, where the A_i are fields and the B_j are bound to their maximal nilideals. If in addition, for every ideal M such that $A - M$ is subdirectly irreducible, $A - M$ satisfies either* d.c.c. *or* a.c.c., *and in particular if A satisfies* d.c.c. *or* a.c.c., *then the B_j are nilpotent.*

References

1. A. A. Albert, *Structure of algebras* (Amer. Math. Soc. Colloquium Publications, *24*, 1939).
2. E. Artin, C. J. Nesbitt and R. M. Thrall, *Rings with minimum condition* (Ann Arbor, 1946).
3. R. Baer, *Radical ideals*, Amer. J. Math., *65* (1943), 537–568.
4. ———, *Kriterien für die Existenz eines Einselements in Ringen*, Math. Z., *56* (1952), 1–17.
5. B. Brown and N. McCoy, *Radicals and subdirect sums*, Amer. J. Math., *69* (1947), 46–58.
6. ———, *The radical of a ring*, Duke Math. J., *15* (1948), 495–499.
7. ———, *The maximal regular ideal of a ring*, Proc. Amer. Math. Soc., *1* (1950), 165–171.
8. M. Hall, *The position of the radical in an algebra*, Trans. Amer. Math. Soc., *48* (1940), 391–404.
9. N. Jacobson, *The radical and semi-simplicity for arbitrary rings*, Amer. J. Math., *67* (1945), 300–320.
10. I. Kaplansky, *Topological rings*, Amer. J. Math., *69* (1947), 153–183.
11. N. McCoy, *Subdirectly irreducible commutative rings*, Duke Math. J., *12* (1945), 381–387.
12. S. Perlis, *A characterization of the radical of an algebra*, Bull. Amer. Math. Soc., *48* (1942), 128–132.
13. ———, *A note on the radical of an ideal*, Bull. Amer. Math. Soc., *53* (1947), 907, abstract 53-9-306.

University of Manitoba

TWO REMARKS ON THE COMMUTATIVITY OF RINGS

I. N. HERSTEIN

In **(1)** and **(2)** we proved that under certain conditions a given ring R must be commutative. The conditions used there were "global" in the sense that they were imposed at once on the relation of a given element to *all* the other elements of the ring R.

In this note we replace these global conditions by "local" ones that relate only to two elements of R at a time. We show that the results of **(1)** and **(2)** carry over to this situation.

In **(1)** we proved that if in a ring R with centre Z, $x^{n(x)} \in Z$ for some integer $n(x) > 1$ for all $x \in R$, then either R is commutative or its commutator ideal is a nil ideal.

The condition $x^{n(x)} \in Z$, of course, means that $x^{n(x)}y = yx^{n(x)}$ for *all* $y \in R$. We prove here

THEOREM 1. *Suppose that R is a ring such that given any two elements $x, y \in R$ then for some integer $n(x, y) > 1$ which depends on both x and y*

$$x^{n(x,y)}y = yx^{n(x,y)}.$$

Then either R is commutative or its commutator ideal is a nil ideal.

Proof. Suppose that R is not commutative. Let $c \neq 0$ be a typical element in the commutator ideal of R. We want to show that c is nilpotent. Since c is in the commutator ideal of R,

$$c = \sum_{i=1}^{m} (a_ib_i - b_ia_i) + \sum_{i=1}^{n} r_i(d_ie_i - e_id_i)$$
$$+ \sum_{i=1}^{p} (f_ig_i - g_if_i)\, s_i + \sum_{i=1}^{q} t_i(h_ik_i - k_ih_i)\, u_i.$$

Let T be the subring of R generated by all the elements a_i, b_i, d_i, e_i, f_i, g_i, h_i, k_i, r_i, s_i, t_i, u_i appearing in the expression for c. Clearly c is a commutator in T. Suppose $\tau \in T$. Then

$$\tau^{n_1} = \tau_1$$

for suitable n_1 commutes with a_1 by the condition imposed on R. Similarly

$$\tau_2 = \tau_1^{n_2} = \tau^{n_1 n_2}$$

commutes with a_2 for some integer n_2; of course τ_2 also commutes with a_1 since τ_1 does. Continuing in this way we arrive at an integer $m > 1$ so that τ^m commutes with all the a's, b's, etc. appearing in the expression for c and which

Received January 10, 1955.

generate T. Thus τ^m commutes with all the elements of T since it commutes with the generators of T; that is, for every $\tau \in T$, $\tau^{n(\tau)}$ is in the centre of T. By **(1)** this means that either T is commutative or its commutator ideal is a nil ideal. Since $c \neq 0$ is in the commutator ideal of T, T is not commutative. So c is nilpotent and the theorem is established.

In **(2)** we proved: let R be a ring with centre Z such that for all $x \in R$ $x^{n(x)} - x \in Z$ for some integer $n(x) > 1$; then R is commutative. The condition $x^{n(x)} - x \in Z$ of course means that

$$(x^{n(x)} - x)y = y(x^{n(x)} - x)$$

for *all* $y \in R$. We localize the condition in the following

THEOREM 2. *If in a ring R for every pair of elements x and y we can find an integer $n(x, y) > 1$ which depends on x and y so that $x^{n(x,y)} - x$ commutes with y, then R is commutative.*

Proof. Let T be the subring of R generated by x and y. Suppose $t \in T$. Thus for some integer $m > 1$, $t_1 = t^m - t$ commutes with x. For some other integer $n > 1$, $t_2 = t^n{}_1 - t_1$ commutes with y. Since t_1 commutes with x, t_2 also commutes with x. Thus t_2 commutes with both x and y, and so with every element in the subring they generate. Thus t_2 is in the centre of T. However

$$t_2 = t_1{}^n - t_1 = (t^m - t)^n - (t^m - t) = -(t^2 p(t) - t)$$

where $p(t)$ is a polynomial with integer coefficients. That is, for every $t \in T$ we can find a polynomial $p(t)$ with integer coefficients so that $t^2 p(t) - t$ is in the centre of T. By the principal result of **(3)** T must be commutative. Since both x and y are in T, $xy = yx$, and so R is commutative.

The main theorem of **(3)** can also be generalized in the same fashion as the other two theorems. We state it without proof,

THEOREM 3. *If for every x and y in R we can find a polynomial $p_{x,y}(t)$ with integer coefficients which depend on x and y such that $x^2 p_{x,y}(x) - x$ commutes with y, then R is commutative.*

REFERENCES

1. I. N. Herstein. "A Theorem on Rings," *Can. J. Math.*, *5* (1953), 238–241.
2. ———, "A Generalization of a Theorem of Jacobson III," *Amer. J. Math.*, *75* (1953), 105–111.
3. ———, "The Structure of a Certain Class of Rings," *Amer. J. Math.*, *75* (1953), 864–871.

University of Pennsylvania

REMARKS ON FINITE GROUPS DEFINED BY GENERATING RELATIONS

ROBERT FRUCHT

The author wishes to make the following correction to his paper *Remarks on finite groups defined by generating relations*, Can. J. Math., 7 (1955), 8–17.

The *duplication principle* stated on p. 8 holds only for those groups $\mathfrak{H}$ which admit an automorphism taking

$$S_1,\ S_1S_2,\ S_1S_2S_3,\ \ldots\ldots,\ S_1S_2S_3\ldots S_k$$

into their respective inverses

$$S_1^{-1},\ (S_1S_2)^{-1},\ (S_1S_2S_3)^{-1},\ \ldots\ldots,\ (S_1S_2S_3\ldots S_k)^{-1}.$$

This fact (which had escaped the author's attention, and for whose discovery he is very much indebted to *Graham Higman* and *B. H. Neumann*) follows immediately from the consideration of that inner automorphism of $\mathfrak{G}$ which takes any element T of $\mathfrak{G}$ into $T_1^{-1}TT_1 = T_1TT_1$; indeed for any m from 2 to $k+1$ we have: $T_1^{-1}(T_1T_m)T_1 = (T_1T_m)^{-1}$, and T_1T_m in $\mathfrak{G}$ corresponds to $S_1S_2S_3\ldots S_{m-1}$ in $\mathfrak{H}$.

If $k = 2$, by using T_2 instead of T_1 the following simpler condition results for a group $\mathfrak{H}$ with 2 generators S_1 and S_2: the *duplication principle* holds if, and only if, $\mathfrak{H}$ admits an automorphism taking S_1 into S_1^{-1}, and S_2 into S_2^{-1}. Hence it does not hold e.g. for the following group $\mathfrak{H}$ of order 21:

$$S_2^{-1}S_1S_2 = S_1^2, \qquad S_2^3 = 1$$

[in permutations: $S_1 = (1\ 2\ 3\ 4\ 5\ 6\ 7)$, $S_2 = (2\ 3\ 5)(4\ 7\ 6)$]; here $\mathfrak{G}$ is easily seen to be of order 6 only instead of 42.

It might be remarked that the *duplication principle* always holds for Abelian groups; whether it holds for a non-Abelian group depends not only on the group, but also on the generators. It holds e.g. for $\mathfrak{D}_6$ (the dihedral group of order 6) when generated by 2 elements (as usually), or by the 3 permutations $S_1 = (1\ 2)$, $S_2 = (1\ 3)$, $S_3 = (1\ 2\ 3)$; but if $\mathfrak{D}_6$ is generated by $S_1 = (1\ 2)$, $S_2 = (1\ 2\ 3)$, $S_3 = (2\ 3)$, the group $\mathfrak{G}$ is of order 4 only, instead of 12.

Technical University Santa Maria, Valparaiso, Chile

THE WEDDERBURN THEOREM

HARRY GOHEEN

Mr. William Scott of the University of Kansas has kindly drawn my attention to an error in my note, *The Wedderburn Theorem*, Can. J. Math., 7 (1955), 60–62. The argument in lines 8–6 from the bottom of page 61 assumes that the division ring generated by an element which leaves a subring invariant and a subfield with the same property, also leaves that subring invariant. This assumption I have not proved, and it invalidates the proof of the theorem.

Iowa State College

OVALS IN A FINITE PROJECTIVE PLANE

BENIAMINO SEGRE

1. Let $\mathfrak{P}$ be a finite projective plane (8, §17), i.e. a projective space of dimension 2 over a Galois field γ. We suppose that γ has *characteristic* $p \neq 2$, hence order $q = p^h$, where p is an odd prime and h is a positive integer. It is well known that every straight line and every non-singular conic of $\mathfrak{P}$ then contains $q+1$ points exactly.

Using the term *oval* to designate any set of $q+1$ distinct points of $\mathfrak{P}$ no three of which are collinear, we shall prove the following theorem, already surmised by Järnefelt and Kustaanheimo (3) (deemed "implausible" in Math. Rev., *14* (1953), p. 1008):

THEOREM I. *If $p \neq 2$, every oval of $\mathfrak{P}$ is a conic (i.e., can be represented by an equation of the second degree).*

This result fills up a gap in the finite congruence axiomatics set up by Kustaanheimo (4), and has important implications if we accept the idea, advanced by Järnefelt (2), of a possible connection between the physical world and the geometry of a finite linear space (cf. also 1, 5, 6, 7).

2. Let $\mathscr{C}$ denote any given oval of $\mathfrak{P}$, and B be an arbitrary point of $\mathscr{C}$. Then $\mathscr{C}$ has a *tangent* at B, uniquely defined as the line of $\mathfrak{P}$ which contains B and no other point of $\mathscr{C}$; moreover, no three tangents of $\mathscr{C}$ meet at a point (7, Theorem 3). We begin by proving

THEOREM II. *Every inscribed triangle of $\mathscr{C}$ and its circumscribed triangle are perspective.*

It is not restrictive to identify the given inscribed triangle with the triangle of reference for homogeneous coordinates (x_1, x_2, x_3):

$$A_1{:}(1,0,0), \quad A_2{:}(0,1,0), \quad A_3{:}(0,0,1);$$

then we may denote by

$$a_1 : x_2 = k_1 x_3, \quad a_2 : x_3 = k_2 x_1, \quad a_3 : x_1 = k_3 x_2$$

the tangents of $\mathscr{C}$ at A_1, A_2, A_3 respectively, where k_1, k_2, k_3 are three non-zero elements of the field γ. If $B{:}(c_1, c_2, c_3)$ is any of the $q-2$ points of $\mathscr{C}$ distinct from A_1, A_2, A_3, then $c_1 c_2 c_3 \neq 0$; moreover, the lines A_1B, A_2B, A_3B have equations of the form

$$x_2 = \lambda_1 x_3, \quad x_3 = \lambda_2 x_1, \quad x_1 = \lambda_3 x_2,$$

Received November 30, 1954.

where the coefficients $\lambda_1, \lambda_2, \lambda_3$ are distinct from k_1, k_2, k_3 respectively, as well as from zero. Since these coefficients are given precisely by

$$\lambda_1 = c_2 c_3^{-1}, \quad \lambda_2 = c_3 c_1^{-1}, \quad \lambda_3 = c_1 c_2^{-1},$$

they satisfy the equation

$$\lambda_1 \lambda_2 \lambda_3 = 1. \tag{1}$$

Conversely, if λ_1 denotes any of the $q-2$ elements of γ distinct from zero and from k_1, the line $x_2 = \lambda_1 x_3$ meets $\mathscr{C}$ at A_1 and at a further point, B say, distinct from A_1, A_2, A_3; hence the coefficients λ_2, λ_3 in the equations $x_3 = \lambda_2 x_1$, $x_1 = \lambda_3 x_2$ of the lines $A_2 B$, $A_3 B$ are functions of λ_1, connected by (1), which take once each of the non-zero values of γ distinct from k_2, k_3 respectively. On multiplying the $q-2$ equations (1) thus obtained, we see that

$$\Pi^3 = k_1 k_2 k_3,$$

where Π denotes the product of the $q-1$ non-zero elements of γ; whence

$$k_1 k_2 k_3 = -1, \tag{2}$$

as it is well known (8, §59) that $\Pi = -1$.

From the equation (2), Theorem II follows at once. In fact the points

$$a_2 \cdot a_3 : \ (k_2, 1, k_2 k_3), \quad a_3 \cdot a_1 : \ (k_3 k_1, k_1, 1), \quad a_1 \cdot a_2 : \ (1, k_1 k_2, k_3)$$

are joined to A_1, A_2, A_3 respectively by the lines:

$$x_3 = k_2 k_3 x_2, \quad x_1 = k_3 k_1 x_3, \quad x_2 = k_1 k_2 x_1;$$

by virtue of (2), these lines concur at the point $K:(1, k_1 k_2, -k_3)$, which is therefore a centre of perspective of the triangles $A_1 A_2 A_3$ and $a_1 a_2 a_3$.

3. We can now prove Theorem I. For this purpose we use the notation of §2, assuming, as it is not restrictive, that K coincides with the unit point $(1, 1, 1)$; this is tantamount to supposing

$$k_1 = k_2 = k_3 = -1.$$

If $B:(c_1, c_2, c_3)$ is any of the $q-2$ points of $\mathscr{C}$ distinct from A_1, A_2, A_3, we denote by

$$b : \ b_1 x_1 + b_2 x_2 + b_3 x_3 = 0$$

the tangent of $\mathscr{C}$ at it. This line contains B, but none of the points A_1, A_2, A_3, $a_2 \cdot a_3, a_3 \cdot a_1, a_1 \cdot a_2$; hence, if we put

$$\beta_1 = b_1 - b_2 - b_3, \quad \beta_2 = -b_1 + b_2 - b_3, \quad \beta_3 = -b_1 - b_2 + b_3,$$

we have

$$b_1 c_1 + b_2 c_2 + b_3 c_3 = 0 \tag{3}$$

and

$$b_1 b_2 b_3 \beta_1 \beta_2 \beta_3 \neq 0. \tag{4}$$

By virtue of Theorem II, the triangles BA_2A_3 and ba_2a_3 are perspective; this—as is immediately seen—is expressed algebraically by the equation

$$\begin{vmatrix} c_3-c_2 & c_1+c_3 & -c_1-c_2 \\ b_1-b_3 & b_2 & 0 \\ b_1-b_2 & 0 & b_3 \end{vmatrix} = 0,$$

i.e., on suppressing the non-zero factor β_1:

$$b_2(c_1+c_2) = b_3(c_1+c_3).$$

Likewise, the consideration of the inscribed triangles BA_3A_1, BA_1A_2 and their circumscribed triangles gives:

$$b_3(c_2+c_3) = b_1(c_2+c_1), \quad b_1(c_3+c_1) = b_2(c_3+c_2).$$

The last three equations imply:

$$b_1 : b_2 : b_3 = (c_2+c_3) : (c_3+c_1) : (c_1+c_2);$$

hence from (3), using also (4) and the hypothesis $p \neq 2$, we deduce the equality

$$c_2c_3 + c_3c_1 + c_1c_2 = 0.$$

This equality means that each of the $q-2$ points B lies on the conic

$$x_2x_3 + x_3x_1 + x_1x_2 = 0.$$

Since this conic obviously contains in addition the three points A_1, A_2, A_3, and its points are precisely $q+1$ in number, thus $\mathscr{C}$ must coincide with it, which proves Theorem I.

4. We remark, in conclusion, that Theorem I does not hold on a finite plane of *characteristic* $p = 2$, if $q > 4$. For, as it is well known, the $q+1$ tangents of a non-singular conic then meet at a point; this point and q of the $q+1$ points of the conic constitute an oval, which, however, is clearly not a conic.

References

1. G. Järnefelt, *A plane geometry with a finite number of elements*, Verröf. Finnischen Geodatischen Inst., *36* (1949), 71–80.
2. ———, *Reflections on a finite approximation to Euclidean geometry*, Ann. Acad. Sci. Fennicae (A, I), *96* (1951), 43 pp.
3. G. Järnefelt and P. Kustaanheimo, *An observation on finite geometries*, Den 11te Skandinaviske Matematikerkongress, Trondheim (1949), 166–182.
4. P. Kustaanheimo, *A note on a finite approximation of the Euclidean plane geometry*, Soc. Sci. Fenn. Comm. Phys. Math., *15*, n. 19 (1950), 11 pp.
5. ———, *On the fundamental prime of a finite world*, Ann. Acad. Sci. Fennicae (A, I), *129* (1952), 7 pp.
6. P. Kustaanheimo and B. Qvist, *On differentiation in Galois fields*, Ann. Acad. Sci. Fennicae (A, I), *137* (1952), 12 pp.
7. B. Qvist, *Some remarks concerning curves of the second degree in a finite plane*, Ann. Acad. Sci. Fennicae (A, I), *134* (1952), 27 pp.
8. B. Segre, *Lezioni di geometria moderna*, vol. *1* (Bologna, 1948).

University of Rome